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, viz., the walker has no memory. In contrast, in the elephant random walk, which was introduced by Schütz and Trimper [11] in 2004, the walker remembers the whole past, and 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.

1 Introduction

In the classical *simple* random walk the steps are equal to plus or minus one and independent—$P(X = 1) = 1 - P(X = -1) = p$, $(0 < p < 1)$. In this model the walker has no memory. This random walk is, in particular, Markovian. Motivated by applications, although 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 [11] in 2004, the name being inspired by the fact that elephants have a very long memory.

The first, more substantial, paper on elephant random walks is, to the best of our knowledge, Bercu's paper [1], in which he proves a number of limit theorems. A main point is that there is a kind of phase transition at the point $p = P(X = 1) = 3/4$, which divides the problem into the diffusive regime, $0 \leq p < 3/4$, the critical regime, $p = 3/4$, and the superdiffusive regime, $3/4 < p \leq 1$, with somewhat different asymptotics.

A main device in his paper is the use of martingale theory due to the observation that a multiplicative scaling of the random walk constitutes a martingale.

Our main interest is the situation in which the elephant has only a limited memory, either that he or she remembers only some distant past, only a recent past or a mixture of both. No paper on exact results and proofs seems to exist, only simulations, in which case a given fraction of the distant/recent past is remembered; [12, 2, 10].

The first task in this direction is to consider the cases when the walker only remembers the first (two) step(s) or only the most recent (previous) step. In particular the latter case involves rather cumbersome computations and we therefore invite the reader(s) to try to push our results further. It should also be mentioned that the paper by Engländer and Volkov [4] is devoted to this latter case, although from a different angle, in that the next step is not generated by flipping a coin, rather by turning it over or not. They have a somewhat different focus, in particular, they consider the case with different p-values in each step.

The cases with limited memory behave very differently mathematically in that some of the walks are still non-markovian others are markovian, but there is no convenient martingale around. Moreover there are no phase transitions in these cases.

A second point concerns the extension of (some of) Bercu's results in [1] from the simple random walk to general sums, that is, to the case when the steps have an arbitrary distribution on the integers.

We begin by defining the various models in Section 2. After some preliminaries in Section 3, some results for general ERW:s are obtained in Section 4. Sections 5 and 6 are devoted to the distant past and Sections 8 and 9 to the recent past, respectively. These "one-sided" memories are then followed up in Sections 10 and 11 where we consider mixed cases, that is, when the memory contains some early steps as well as some recent ones, after which we shortly discuss some different models. We close with a section containing some questions and remarks. For easier reading we collect some of the somewhat 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

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

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

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

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.

Our main aim is to extend these results to the case when the elephant has only 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 remain true (suitably modified) also when the steps of the ERW:s 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, viz.,

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

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

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

where $P(K=1) = P(K=2) = 1/2$.

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

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

We begin, throughout, by assuming that $X_1 = 1$, and generalize our findings in this setting (for simplicity) to the case $r = p$. We denote our partial sums with 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, viz., the corresponding walk equals the mirrored image in the time axis. This implies that the mean after n steps equals $-E(T_n)$, but the dynamics being the same, 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 11 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, and $p = 0$ the the case of alternating summands.

3   Some auxiliary material

For easier access of the arguments below we shortly 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 $U_n \overset{d}{\to} U$ as $n \to \infty$, then $U_n V \overset{d}{\to} UV$ as $n \to \infty$.*

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

$$
\begin{aligned}
\varphi_{U_n V}(t) &= E \exp\{itU_n V\} = E\Big(E(\exp\{itU_n V\} \mid V)\Big) = E\varphi_{U_n}(tv) \to E\varphi_U(tv) \\
&= E\Big(E(\exp\{itUV\} \mid V)\Big) = E \exp\{itUV\}) = \varphi_{UV}(t).
\end{aligned}
$$

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

3.2   Conditioning in case of a restricted memory

Let $\{S_n, n \geq 1\}$ be an ERW, let $\{\mathcal{F}_n, n \geq 1\}$ denote the σ-algebras generated by the memory of the elephant and let $\mathcal{G}_n = \sigma\{X_1, X_2, \ldots, X_n\}$ stand for the full memory. We already know from (2.1) above that $E(X_{n+1} \mid \mathcal{G}_n) = (2p-1)S_n/n$. Our aim is to establish analogs when the elephant has a restricted memory, that is, analogs for $E(X_{n+1} \mid \mathcal{F}_n)$.

Toward the end, let $I_n = \{i \leq n : i \in \mathfrak{M}\}$, where $\mathfrak{M} =$ the memory of the elephant. Then,

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

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, $I_n = \{n\}$ the elephant only remembers the most recent step, and $I_n = \{1, n\}$ means that he/she only remembers the first and the most recent steps; these are two cases that will be considered in the sequel. In these cases (3.1) states that

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

respectively.

The next problem is when we condition on steps that are not contained in the memory. In words, if they do not, the elephant does not remember them, and, hence, cannot choose among them in a following step. More precisely, mathematically $\mathfrak{M}$ is defined as those steps in the past on which the elephant bases the next step. Technically, let $I \subset \{1, 2, \ldots, n\}$ be an arbitrary set of indices, such that $I \cap I_n = \emptyset$. Then

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

It follows, in particular, that

$$
E(X_{n+1} \mid \mathcal{G}_n) = (2p-1)\frac{\sum_{i \in I_n} X_i}{|I_n|}, \tag{3.3}
$$

and that

$$E(S_n X_{n+1} \mid \mathcal{G}_n) = S_n E\big(X_{n+1} \mid \mathcal{G}_n\big) = S_n(2p-1)\frac{\sum_{i \in I_n} X_i}{|I_n|}. \tag{3.4}$$

This, and the fact that $X_{n+1}^2 = 1$, will be useful several times for the computation of second moments as follows:

$$
\begin{aligned}
E(S_{n+1}^2) &= E(S_n + X_{n+1})^2 = E(S_n^2) + 2E(S_n X_{n+1}) + E(X_{n+1}^2) \\
&= E(S_n^2) + \frac{2(2p-1)}{|I_n|}E\Big(S_n \sum_{i \in I_n} X_i\Big) + 1. \tag{3.5}
\end{aligned}
$$

3.3   Difference equations

In the proofs we use several difference equations. For convenience and easy reference we summarize here some well-known facts about linear difference equations that are used on and off.

Proposition 3.2 (i) *Consider the first order equation*

$$x_{n+1} = a\,x_n + b_n, \quad for \quad n \geq 1, \quad with \quad x_1^* \quad given.$$

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

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

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

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

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

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

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

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

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

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

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 $y_n = d/(1 - a - b)$.

3.4   Some notation

We use the standard $\delta_a(x)$ to denote the distribution function with a jump of height one at a. 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 $\widetilde{S}_n/a_n \overset{a.s.}{\to} Z$ 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 $R\widetilde{S}_n/a_n \overset{a.s.}{\to} RZ$ 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 [1] 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 constructeded 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 results of [1] if R is a coin-tossing random variable), holds for $S_n = R\widetilde{S}_n$:

Theorem 4.1   (a)   *For* $0 < p < 3/4$, $\dfrac{S_n}{n} \overset{a.s.}{\to} 0$ *as* $n \to \infty$;

(b)   *For* $p = 3/4$, $\dfrac{S_n}{\sqrt{n}\log n} \overset{a.s.}{\to} 0$ *as* $n \to \infty$;

(c)   *For* $3/4 < p < 1$, $\dfrac{S_n}{n^{2p-1}} \overset{a.s.}{\to} RL$ *as* $n \to \infty$,

where L is a non-dgenerate 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*

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

Moreover, if $E(R^2) < \infty$, then $E(S_n/\sqrt{n}) \to 0$ and $E((S_n/\sqrt{n})^2) \to E(R^2)/(3 - 4p)$ as $n \to \infty$.

PROOF. As R and S_n are independent we find that

$$
\begin{aligned}
P\Big(\frac{S_n}{\sqrt{n}} \leq x\Big) &= \int_{\mathbb{R}} P\Big(R\frac{\widetilde{S}_n}{\sqrt{n}} \leq x \mid R\Big) dF_R(t) \\
&= \int_{\mathbb{R}} P\Big(t\frac{\widetilde{S}_n}{\sqrt{n}} \leq x\Big) dF_R(t) \\
&= \int_{(-\infty,0)} P\Big(\frac{\widetilde{S}_n}{\sqrt{n}} \geq x/t\Big) dF_R(t) + \int_{(0,\infty)} P\Big(\frac{\widetilde{S}_n}{\sqrt{n}} \leq x/t\Big) dF_R(t) \\
&\quad + P(R = 0) \cdot \delta_{[0,\infty)}(x) \\
&\to \int_{(-\infty,0)} \big(1 - \mathcal{N}_{0,\frac{1}{3-4p}}(x/t)\big) dF_R(t) + \int_{(0,\infty)} \mathcal{N}_{0,\frac{1}{3-4p}}(x/t)\, dF_R(t) \\
&\quad + P(R = 0) \cdot \delta_{[0,\infty)}(x),
\end{aligned}
$$

by dominated convergence which yields the desired result.

The second part is immediate, since R is independent of everything else.   □

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 [1].

REMARK 4.2 For the critical case, $p = 3/4$ one similarly obtains, using [1], Theorem 3.6, that

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

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

5   Remembering only the distant past 1

This turns out as being the easiest case, since convenient independence is inherent. We begin by assuming that the elephant only remembers the first step, i.e., that $\mathcal{F}_n = \sigma\{X_1\}$, and begin with the assumption that $X_1 = 1$ (recall that partial sums are denoted with the letter T). Then,

$$E(X_{n+1} \mid \mathcal{F}_n) = E(X_{n+1} \mid X_1) = (2p - 1) \cdot 1 = E(X_{n+1}) \quad \text{for all} \quad n \geq 1,$$

and, hence,
$$E(T_{n+1} \mid \mathcal{F}_n) = 1 + n(2p-1) = E(T_{n+1}).$$

Moreover, applying (3.5) to T_n we find that
$$\begin{aligned} E(T_{n+1}^2) &= E(T_n^2) + 2\,(2p-1)E(T_n \cdot 1) + 1 \\ &= E(T_n^2) + 2(2p-1)\big(1 + (n-1)(2p-1)\big) + 1 \\ &= E(T_n^2) + 2(2p-1)^2 n + 4(2p-1)(1-p) + 1, \end{aligned}$$

which, after telescoping, yields
$$E(T_{n+1}^2) = 1 + (2p-1)^2 n(n+1) + \big(4(2p-1)(1-p) + 1\big)n,$$

and, finally,
$$\mathrm{Var}\,(T_{n+1}) = \big(1 - (2p-1)^2)\big)n = 4p(1-p)n.$$

A completely analogous calculation for characteristic functions, with an eye on (3.2), shows that
$$\begin{aligned} \varphi_{T_{n+1}}(t) &= E\Big(E\big(e^{it(T_n + X_{n+1})} \mid \mathcal{G}_n\big)\Big) = E\Big(e^{itT_n} \cdot E\big(e^{itX_{n+1}} \mid \mathcal{G}_n\big)\Big) \\ &= E\Big(e^{itT_n} \cdot E\big(e^{itX_{n+1}} \mid X_1\big)\Big) = E\Big(e^{itT_n} \cdot (pe^{it} + (1-p)e^{-it})\Big) \\ &= \varphi_{T_n}(t) \cdot \big(pe^{it} + (1-p)e^{-it}\big), \end{aligned}$$

after which telescoping tells us that
$$\varphi_{T_{n+1}}(t) = \big(pe^{it} + (1-p)e^{-it}\big)^n \cdot e^{it},$$

after which a standard computation shows that
$$\varphi_{(T_n - n(2p-1))/\sqrt{n}}(t) \to \exp\{-4p(1-p)t^2\} \quad \text{as} \quad n \to \infty.$$

Next we note that the computations so far prove that the increments are *uncorrelated*, suggesting independence ... In fact, recalling that we have assumed that $X_1 = 1$, we have, setting $\alpha_k = 1$ if $X_k = +X_1$, for $k \geq 2$, and 0 otherwise,
$$\begin{aligned} P(X_i = X_j = 1) &= P(\alpha_i = \alpha_j = 1 - p) = P(\alpha_i = 1 - p) \cdot P(\alpha_j = 1 - p) \\ &= (1-p)^2 = P(X_i = 1) \cdot P(X_j = 1), \end{aligned}$$

for $i, j \geq 2$ and different.

This means that the ERW coincides with the classical simple random walk, except for the fact that the first step is always equal to one. This is—after some thinking—rather obvious, because (in the language of [4]) we might interpret X_1 as a coin that we either flip or not before each new step. Hence we obtain:

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

If, on the other hand, the first step is equal to -1, then, by symmetry, $E(T_{n+1}) = -n(2p-1)-1$, the variance remains the same (recall the discussion toward the end of Section 2), and, again, by symmetry, $T_{n+1} + n(2p-1)$ normalized by $\sqrt{n}$ is asymptotically normal and the SLLN and the LIL do hold again.

As a consequence, assuming that X_1 is a coin-tossing random variable, we are (asymptotically) confronted with two normal distributions, one for each of the two portions of the probability space. In fact, if we imagine the situation that $r = P(X_1 = +1)$ is close to zero or one it is rather apparent how the very first step determines along which branch it will evolve.

One also notes, more formally, that $E(X_{n+1} \mid \mathcal{F}_n) = (2p-1)E(X_1)$, so that $E(X_{n+1}) = (2p-1)^2$, implying that $\mathrm{Var}\,(S_n) = \mathcal{O}(n^2)$ (and not of order n) as $n \to \infty$. Thus, an ordinary CLT is not valid, with the exception that if $p = 1/2$ the two "branches" determined by the first step collaps (asymptotically) into one, and we are ultimately faced with a classical simple symmetric random walk.

Hence, the following limit result is always available in the general case:

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

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

(b) $\qquad E(S_n/n) \to (2p-1)^2 \quad \text{and} \quad \text{Var}\,(S_n/n) \to 4p(1-p)(2p-1)^2 \quad \text{as} \quad n \to \infty.$

PROOF OF (A). If $X_1 = \pm 1$ we know from above that $E(T_n) = \pm(1+(n-1)(2p-1))$, and that $\text{Var}\,(T_n) = 4p(1-p)(n-1)$. This tells us that, $\frac{T_n}{n} \xrightarrow{P} \pm(2p-1)$ as $n \to \infty$. The conclusion follows.
PROOF OF (B). Immediate. $\hfill\square$

REMARK 5.1 (i) 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 $\pm(2p-1)$.

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

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

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

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

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

PROOF OF (A). We use the fact that

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

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.

PROOF OF (B) AND (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 $S_n - n(2p-1)X_1$. Similarly on Ω_2. Combining them yields the desired result. $\hfill\square$

REMARK 5.2 The strong law can also be formulated with a random RHS:

$$\frac{S_n}{n} \xrightarrow{a.s.} (2p-1)X_1 \qquad \text{as} \quad n \to \infty.$$

REMARK 5.3 If X_1 is a general random variable with distribution F having no mass at zero, then

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

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

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

$$\frac{S_n}{n} \xrightarrow{a.s.} 0 \text{ as } n \to \infty, \qquad \frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \text{ as } n \to \infty, \qquad \underset{n \to \infty}{\limsup}\,(\underset{n \to \infty}{\liminf})\, \frac{S_n}{\sqrt{2n\log\log n}} = 1;(-1) \qquad a.s.$$

6   Remembering only the distant past 2

In this section we begin by assuming that the elephant only remembers the first two steps, so that $\mathcal{F}_n = \sigma\{X_1, X_2\}$, and suppose that $X_1 = X_2 = 1$. Then, for $n \geq 2$,

$$E(X_{n+1} \mid \mathcal{F}_n) = E(X_{n+1} \mid X_1, X_2) = (2p-1) \cdot \frac{1+1}{2} = (2p-1) = E(X_{n+1}) \qquad (6.1)$$

for all $n \geq 2$, and, hence,

$$E(T_{n+1} \mid \mathcal{F}_n) = 2 + (n-1)(2p-1) = n(2p-1) + 3 - 2p = E(T_{n+1}),$$

(since $E(T_2) = 2$). Extending the idea from the previous section that the walk evolves as an ordinary simple random walk beginning at the third step, a natural guess is that

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

To see this we first observe that $\text{Var}\,(T_3) = \text{Var}\,(X_3) = 4p(1-p)$, that is, the formula is correct for $n = 2$. Assuming it is correct for $n - 1$ we have

$$
\begin{aligned}
\text{Var}\,(T_{n+1}) &= \text{Var}\,(T_n + X_{n+1}) = \text{Var}\,(T_n) + 2\text{Cov}\,(T_n, X_{n+1}) + \text{Var}\,(X_{n+1}) \\
&= 4p(1-p)(n-2) + 0 + 4p(1-p) = 4p(1-p)(n-1),
\end{aligned}
$$

since by (3.4) and the fact that $X_1 = X_2 = 1$,

$$\text{Cov}\,(T_n, X_{n+1}) = E(T_n X_{n+1}) - E(T_n)E(X_{n+1}) = E(T_n(2p-1)) - E(T_n)(2p-1) = 0. \qquad (6.2)$$

Next, by modifying the computations involving the characteristic function from Section 5, we obtain

$$
\begin{aligned}
\varphi_{T_{n+1}}(t) &= E\Big(E\big(e^{it(T_n + X_{n+1})} \mid \mathcal{G}_n\big)\Big) = E\Big(e^{itT_n} \cdot E\big(e^{itX_{n+1}} \mid X_1, X_2\big)\Big) \\
&= E\Big(e^{itT_n} \cdot \big(pe^{it} + (1-p)e^{-it}\big)\Big) = \varphi_{T_n}(t) \cdot \big(pe^{it} + (1-p)e^{-it}\big).
\end{aligned}
$$

By continuing as before one obtains, after proper centering, a limiting normal distribution for these initial X-values. Similarly for the other ones. But, only for each "branch" separately. One can also ascertain that the variance is not linear if we assume random beginnings. Except, as before, when $p = 1/2$ and 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) $\qquad \dfrac{S_n}{n} \xrightarrow{d} \begin{cases} 2p-1, & \text{with probability} \quad p^2, \\ 0, & \text{with probability} \quad 1-p, \\ -(2p-1), & \text{with probability} \quad p(1-p), \end{cases} \quad \text{as} \quad n \to \infty;$

(b) $\qquad E(S_n/n) \to p(2p-1)^2 \quad \text{and that} \quad \text{Var}\,(S_n/n) \to p(1-p)(2p-1)^2\big(4p^2 + 1\big).$

PROOF OF (A). If $X_1 = X_2 = \pm 1$ we know from above that $E(T_n) = \pm(n(2p-1) + 3 - 2p)$, and that $\text{Var}\,(T_n) = 4p(1-p)(n-2)$. Moreover, $E(T_n) = 0$ whenever X_1 and X_2 have different signs. The variance remains the same (with $p = 1/2$). This, together with the fact that $P(X_1 = X_2 = 1) = p^2$, $P(X_1 = X_2 = -1) = (1-p)p$, and $P(X_1 \neq X_2) = p(1-p) + (1-p)^2 = 1 - p$ helps us to finish the proof of the first part. Part (b) follows. $\qquad \square$

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) Mimicing Remark 5.1 we may rewrite the conclusion of the theorem is

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

Once again random normalization produces further limit results:

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

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

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

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

where $\sigma(X_1,X_2) = \begin{cases} 4p(1-p), & for \quad \omega \in \{\omega \in \Omega : X_1(\omega)\cdot X_2(\omega) = 1\}, \\ 1, & otherwise. \end{cases}$

PROOF OF (A). Conditioning on the value of $(X_1+X_2)/2$ we obtain

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

$$+ P\Big(\frac{S_n - n(2p-1)(X_1+X_2)/2}{\sqrt{n}} \le x \mid X_1 = X_2 = -1\Big)\cdot p(1-p)$$

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

$$= P\Big(\frac{T_n - n(2p-1)}{\sqrt{n}} \le x\Big)\cdot p^2 + P\Big(\frac{-T_n + n(2p-1)}{\sqrt{n}} \le x\Big)\cdot p(1-p)$$

$$+ P\Big(\frac{T_n}{\sqrt{n}} \le x \mid X_1 + X_2 = 0\Big)\cdot(1-p)$$

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

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

Parts (b) and (c) follow along the lines of the proof of Theorem 5.1. $\qquad\square$

7 The distant past; higher order

If one remembers the first m random variables for some $m \in \mathbf{N}$, the following obvious extension of the above results emerges.

Theorem 7.1 *For* $q_k = P(S_m = m - 2k)$, $r_k = \big((m-k)p + k(1-p)\big)/m$, *and* $p_k = (m-2k)(2p-1)/m$, *where* $0 \le k \le m$ *and* $m \in \mathbf{N}$,

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

and

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

PROOF. As before we write

$$P\Big(\frac{S_n}{n} \le x\Big) = P\Big(\frac{S_n}{n} \le x \mid X_1 = 1,\dots,X_m = 1\Big)\cdot P(X_1 = 1,\dots,X_m = 1)$$

$$+ P\Big(\frac{S_n}{n} \le x \,\big|\, \text{exactly one} - 1 \text{ among the first } m \ X\text{:es}\Big)$$

$$\times P(\text{exactly one} - 1 \text{ among the first } m \ X\text{:es})$$

$$+ \cdots + P\Big(\frac{S_n}{n} \le x \mid X_1 = -1,\dots,X_m = -1\Big)\cdot P(X_1 = -1,\dots,X_m = -1)$$

$$= P\Big(\frac{S_n}{n} \le x \mid X_1 = 1,\dots,X_m = 1\Big)\cdot P(S_m = m)$$

$$+ P\Big(\frac{S_n}{n} \le x \,\big|\, \text{exactly one} - 1 \text{ among the first } m \ X\text{:es}\Big)\cdot P(S_m = m - 2)$$

$$+ \cdots + P\Big(\frac{S_n}{n} \le x \mid X_1 = -1,\dots,X_s = -1\Big)\cdot P(S_m = -m),$$

and observe that in each conditional case we have a random walk with the appropriate success probabilities, i.e., for $S_m = m - 2k$ the sucess probability is $r_k = \big((m-k)p + k(1-p)\big)/m$, and, hence, the expectation is $p_k = 2r_k - 1 = (m - 2k)(2p - 1)/m$.

REMARK 7.1 (i) The probabilities at the jumps are relatively complicated and therefore not expressed in detail, but $q_0 = p^m$ and $q_m = (1-p)p^{m-1}$.

(ii) A more detailed analysis shows that the probability mass of the limit distribution of S_n/n concentrates near zero as m increases.

(iii) One easily checks that the variance for each "branch" equals $4p(1-p)(n-m)$, which, in turn, is dominated by $\leq 4p(1-p)n$, which, consequently, tells us that the analog of Theorems 5.1 and 6.1 holds.

(iv) Once again, the case $p = 1/2$ is special as described in the two previous sections.   □

8   Remembering only the recent past 1

This situation is much more complex, because, even though one remembers only recent steps, the path depends on the whole history so far (some remarks on that will be given in Subsection 12.3). Once again we begin by assuming that the elephant only remembers the very last step, which means that $\mathcal{F}_n = \sigma\{X_n\}$. This setting is reminiscent of [4], where one turns over a coin instead of tossing it. The main focus there, however, is on different p-values at each step and, e.g., how this may affect phase transitions and behavior at critical values.

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

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

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

$$E(X_{n+1}) = (2p-1)^n E(X_1) = (2p-1)^n, \tag{8.1}$$

and

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

For the second moment we have, by (3.5) and (3.2),

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

For the middle term we obtain by (3.2),

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

which in turn, after iteration, yields

$$E(T_n X_n) = 1 + \sum_{k=1}^{n-1}(2p-1)^k = \frac{1 - (2p-1)^n}{2(1-p)} \,.$$

Now we can calculate the second moment:

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

By telescoping we obtain

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

which implies the following formula for the asymptotic variance:

$$\text{Var}\,(T_{n+1}) = \frac{np}{1-p} + \mathcal{O}(1) \quad \text{as} \quad n \to \infty. \tag{8.2}$$

Noticing that $S_n = X_1 T_n$ and that $X_1 = \pm 1$, a glance at (8.1) and (8.2) shows that $\frac{T_n}{n} \overset{p}{\to} 0$ and that $\frac{S_n}{n} \overset{p}{\to} 0$ as $n \to \infty$, suggesting the following result:

Theorem 8.1 *For $X_1 = \pm 1$,*

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

Our next task is to apply the method of moments in order to prove that this is indeed true. We thus wish to prove that

$$E\left(\frac{T_n}{n}\right)^m \to \mu_m = \begin{cases} \dfrac{m!}{2^{m/2}(m/2)!} \cdot \left(\dfrac{p}{1-p}\right)^{m/2}, & \text{if } m \text{ is even,} \\ 0, & \text{if } m \text{ is odd.} \end{cases} \tag{8.3}$$

This amounts to lengthy computations of various higher order mixed moments. The reason for this is that higher order moments of T_n can be expressed as linear combinations of lower order moments of T_n and X_n with the aid of the binomial theorem.

Convergence of mean and variance has already been established above. For higher order moments we use induction.

Throughout in the following, $C(p,m)$, with our without an index, are numerical constants which may differ from line to line and $E(R_n(p,m))$ are quantities of smaller order than the leading term.

Lemma 8.1 *For $m \geq 1$ we have, as $n \to \infty$,*

$$\begin{aligned} E\big((T_n)^{2m-1}X_n\big) &= (2p-1)E\big((T_{n-1})^{2m-1}X_{n-1}\big) + (2m-1)E(T_{n-1})^{2m-2} + E(R_n(p,m)) \\ &\sim \frac{2m-1}{2(1-p)} \cdot E(T_n)^{2m-2}\,; \end{aligned} \tag{8.4}$$

$$\begin{aligned} E\big((T_n)^{2m}X_n\big) &= (2p-1)E\big((T_{n-1})^{2m}X_{n-1}\big) + 2m \cdot E(T_{n-1})^{2m-1} + \\ &\qquad + m(2m-1)\,(2p-1)E\big((T_{n-1})^{2m-2}X_{n-1}\big) + E(R_n(p,m)) \\ &\sim C_1(p,m) \cdot n^{m-1}\,; \end{aligned} \tag{8.5}$$

$$\begin{aligned} E(T_{n+1})^{2m} &= E((T_n)^{2m}) + 2m(2p-1)\,E((T_n)^{2m-1}X_n) + \\ &\qquad + m(2m-1)E((T_n)^{2m-2}) + R_n(p,m) \\ &= E((T_n)^{2m}) + \frac{m(2m-1)\,(2p-1)}{1-p}E((T_n)^{2m-2}) + E(R_n(p,m)) \\ &\sim \frac{(2m)!}{2^m m!}\left(\frac{p}{1-p}\right)^m (n+1)^m\,; \end{aligned} \tag{8.6}$$

$$\begin{aligned} E(T_{n+1})^{2m+1} &= E((T_n)^{2m+1}) + (2m+1)(2p-1)E((T_n)^{2m}X_n) \\ &\qquad + (2m+1)(m-1)E((T_n)^{2m-1}) + E(R_n(p,m)) \\ &\sim C_2(p,m) \cdot (n+1)^m\,, \end{aligned} \tag{8.7}$$

where $E(R_n(p,m))$ denotes individual remainder terms.

The proof of the lemma amounts to extending the above computations for mean and variance to higher order variants and is deferred to the Appendix, Subsection A.1.

PROOF OF THEOREM 8.1. As already mentioned, the proof exploits the method of moments. For $X_1 = 1$ the lemma tells us that

$$E\big((T_n/\sqrt{n})^{2m}\big) \to \frac{(2m)!}{2^m m!}\left(\frac{p}{1-p}\right)^m \quad \text{and that} \quad E\big((T_n/\sqrt{n})^{2m+1}\big) \to 0 \quad \text{as} \quad n \to \infty,$$

which verifies (8.3). For $X_1 = -1$ we recall from the end of Section 3 that even moments remain the same and that odd moments are the same except for a change of sign, which yields the same conclusion. The limit result for S_n then follows as in Theorem 4.2. □

REMARK 8.1 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 also follows from a CLT for Markov chains, see, e.g., Corollary 5 of [8] (cf. also [7], Theorem 19.1.) □

The Markov property also provides a strong law.

Theorem 8.2 *We have*

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

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 in [3] yields the conclusion. □

9 Remembering only the recent past 2

In this section we assume that the elephant remembers the two most recent steps, that is, at time n the next step is based on the steps X_n and X_{n-1}. The computations are as before, although more elaborate. We have, as always, $X_1 = 1$, $E(X_2 \mid \mathcal{F}_1) = (2p-1)X_1$,

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

and, for $n \geq 3$,

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

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

Lemma 9.1 *As $n \to \infty$,*

$$
\begin{aligned}
E(X_n) &\to 0; \\
E(S_n) &\to \frac{(2p-1)(2p+1)}{4(1-p)}; \\
\mathrm{Var}\,(S_n/\sqrt{n}) &\to 1 + \frac{(2p-1)(5-2p)}{2(1-p)(3-2p)} = \sigma_2^2.
\end{aligned}
$$

The expectation of X_n tends to zero geometrically fast.

REMARK 9.1 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 9.1 *We have*

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

PROOF. The sequence $\{X_n, n \geq 1\}$ now forms a Markov chain of order two. Theorem 6.1 in [3] yields the strong law, and the results in [5], Section 3, or [6], combined with Corollary 5 of [8], yield the asymptotic normality with the moments as calculated above. □

REMARK 9.2 If we suppose that the elephant remembers a fixed but finite number, k say, 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 and use e.g. [6]). □

10 Remembering the distant as well as the recent past 1

Next we consider the case when the elephant has a clear memory of the early steps as well as the very recent ones.

One can think of a(n old) person who remembers the early childhood and events from the last few days but nothing in between. The most elementary case is $\mathcal{F}_n = \sigma\{X_1, X_n\}$, for all $n \geq 2$. Following the approach of earlier variants we begin by assuming that $X_1 = 1$. Then, for $n \geq 2$,

$$E(X_2) = E\big(E(X_2 \mid X_1)\big) = E\big((2p-1)X_1\big) = 2p - 1,$$

and

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

Exploiting Proposition 3.2(i) we obtain, for $n \geq 1$,

$$E(X_n) = \frac{2p-1}{3-2p} + \left(\frac{2p-1}{2}\right)^{n-1} \cdot \frac{4(1-p)}{3-2p}, \tag{10.1}$$

and, hence, that

$$
\begin{aligned}
E(T_n) &= 1 + (2p-1) + (n-2)\frac{2p-1}{3-2p} + \frac{4(1-p)}{3-2p}\sum_{k=1}^{n}\left(\frac{2p-1}{2}\right)^{k-1} \\
&= n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2} + o(1) \quad \text{as} \quad n \to \infty.
\end{aligned} \tag{10.2}
$$

Next we note that $E(T_1^2) = 1$, and, by (3.5), that, for $n \geq 1$,

$$E(T_{n+1}^2) = E(T_n^2) + (2p-1)E(T_n) + (2p-1)E(T_nX_n) + 1. \tag{10.3}$$

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.7)),

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

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

$$
\begin{aligned}
\text{Var}\,(T_n) &= n^2 \cdot \frac{(2p-1)^2}{(3-2p)^2} + n \cdot \left(1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 40p + 35)\right) \\
&\quad - \left(n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2}\right)^2 + o(n) \\
&= n \cdot \sigma_T^2 + o(n) \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{10.4}
$$

where

$$\sigma_T^2 = 1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 24p + 19). \tag{10.5}$$

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

$$\frac{T_n}{n} \xrightarrow{p} \frac{2p-1}{3-2p} \quad \text{as} \quad n \to \infty. \tag{10.6}$$

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 $\{T_n, \, n \geq 1\}$ 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 [8] (cf. also [7], Theorem 19.1) to conclude that $T_n - E(T_n)$ is asymptotically normal with mean zero and variance $\sigma_T^2 n$, with σ_T^2 as defined in (10.5), which, in view of (10.2), establishes that

$$\frac{T_n - \frac{2p-1}{3-2p}n}{\sigma_T\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as} \quad n \to \infty. \tag{10.7}$$

An appeal to the disussion at the end of Section 2 now allows us to conclude that

$$
\begin{aligned}
E(S_n) &= pE(T_n) + (1-p)E(-T_n) = (2p-1)E(T_n), \\
E(S_n^2) &= E(T_n^2), \\
\text{Var}\,(S_n) &= E(T_n)^2 - (2p-1)^2(E(T_n))^2,
\end{aligned}
$$

which tells us that

$$E(S_n/n) \to \frac{(2p-1)^2}{3-2p} \quad \text{and} \quad \text{Var}\,(S_n/n) \to \sigma_S^2 = 4p(1-p)\frac{(2p-1)^2}{(3-2p)^2} \quad \text{as} \quad n \to \infty. \tag{10.8}$$

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

Theorem 10.1 *We have*

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

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

REMARK 10.1 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. $\qquad\square$

Finally, by combining (10.7) with the obvious analog for the case $X_1 = -1$, asymptotic normality follows with a *random* centering:

Theorem 10.2 *We have*

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

PROOF. We first note that it follows from the discussion following (10.7) that the CLT there remains true when $X_1 = -1$ with a $+$ replacing the $-$ in the numerator. We thus may 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. $\qquad\square$

11 Remembering the recent as well as the distant past 2

In this section we extend the previous one in that we assume that $\mathcal{F}_n = \sigma\{X_1, X_2, X_n\}$, for all $n \ge 3$. Following the approach of earlier variants we begin by assuming that $X_1 = X_2 = 1$. Then $E(X_1) = E(X_2) = 1$, $E(X_3) = (2p-1)$ and, for $n \ge 3$,

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

Exploiting Proposition 3.2(i) yields

$$
E(X_n) = \frac{2p-1}{2-p} + \left(\frac{2p-1}{3}\right)^{n-1}\left(1 - \frac{2p-1}{2-p}\right) = \frac{2p-1}{2-p} + \frac{3(1-p)}{2-p}\left(\frac{2p-1}{3}\right)^{n-2}, \qquad (11.1)
$$

and, hence,

$$
\begin{aligned}
E(T_n) &= 1 + 1 + (2p-1) + (n-3) \cdot \frac{2p-1}{2-p} + \frac{3(1-p)}{2-p} \sum_{k=4}^{n}\left(\frac{2p-1}{3}\right)^{k-2} \\
&= n \cdot \frac{2p-1}{2-p} + \frac{3(1-p)(7-2p)}{2(2-p)^2} + o(1) \quad as \quad n \to \infty. \qquad (11.2)
\end{aligned}
$$

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

$$
E(T_{n+1}^2) = E(T_n^2) + 2E(T_n X_{n+1}) + 1. \qquad (11.3)
$$

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

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

The variance, finally, turns out as

$$\text{Var}\,(T_n) = n \cdot \sigma_T^2 + o(n), \quad \text{where} \quad \sigma_T^2 = 1 + \frac{(2p-1)}{(2-p)^3} \cdot (5 - 5p - p^2). \tag{11.4}$$

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

$$\frac{T_n}{n} \xrightarrow{p} \frac{2p-1}{2-p} \quad \text{as} \quad n \to \infty. \tag{11.5}$$

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 $E(-T_n)$ and the second moment equals $E(T_n^2)$. 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,

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

and, for $n \geq 3$,

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

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

$$E(T_n) = 0. \tag{11.6}$$

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

$$E(T_n^2) = n \cdot \frac{1+p}{2-p} + o(n) = \text{Var}\,(T_{n+1}) \quad \text{as} \quad n \to \infty,$$

where the last equality is due to the fact that $E(T_n) = 0$. The weak law now runs slightly differently, in that

$$\frac{T_n}{n} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{11.7}$$

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 10, in that

$$\frac{T_n \pm \frac{2p-1}{2-p}n}{\sigma_T \sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as} \quad n \to \infty, \tag{11.8}$$

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

$$\frac{T_n}{\sqrt{n \cdot \frac{1+p}{2-p}}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as} \quad n \to \infty, \tag{11.9}$$

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,

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

Collecting the various pieces tells us that

$$E(S_n/n) \to \frac{p(2p-1)^2}{2-p} \quad \text{and} \quad \text{Var}\,(S_n/n) \to \frac{p(1-p)(2p-1)^2(4p^2+1)}{(2-p)^2} \quad \text{as} \quad n \to \infty.$$

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

Theorem 11.1 *We have*

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

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

We finally wish to combine the asymptotic normality for 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 10. However, there is a problem with the variance. Namely, in Theorem 10.2 both cases had the same variance, whereas here 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 11.2 *We have*

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

with σ_T^2 as given in (11.4).

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.   □

12   Miscellania

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

12.1   More on restricted memories

(i) The next logical step would be to check the case when $\mathcal{F}_n = \sigma\{X_1, X_{n-1}, X_n\}$. By modifying the computations in Appendix A.2, setting $a = \frac{2p-1}{3}$ and $d = 3a^2$, we find that

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

after which Proposition 3.2(iv), and a glance at the computations in Appendix A.2, tell us that

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

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

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

If $\mathcal{F}_n = \sigma\{X_1, X_2, X_{n-1}, X_n\}$, then, with $a = \frac{2p-1}{4}$ and $d = 2p(2p-1)^2/4$, one similarly obtains that

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

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) A more subtle case is when the number of memory steps depends on n, such as $\log n$ or $\sqrt{n}$.

(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 $p_k = P(X_{k+1} = 1)$, $V_n = X_{k+1} + \ldots X_n$, and $\mathcal{H}_n = \sigma\{X_{k+1}, \ldots X_n\}$, and let $n \geq k+1$. Then

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

where $\gamma_n = (n-3+2p_k)/(n-2)$. With

$$\tilde{a}_n = \prod_{k=3}^{n} \tilde{\gamma}_k^{-1} = \frac{\Gamma(n-1)\,\Gamma(2p_k)}{\Gamma(n-2+2p_k)}$$

one can, as in [1], show that $\tilde{a}_n V_n$ is a martingale. From the same paper it follows, provided that $0 < p_k < 3/4$, that

$$\frac{V_n}{\sqrt{n-2}} \xrightarrow{d} \mathcal{N}_{0,1/(4-3p_k)} \quad \text{as} \quad n \to \infty,$$

which implies that

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

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 $\alpha\sqrt{n}$ steps and or the last $\beta\sqrt{n}$ steps for some $\alpha, \beta > 0$;

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

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

12.2 Phase transition

The results of Bercu [1] show that for the full memory one has a phase transition at $p = 3/4$. There is no such thing in our results. An obvious, as well as interesting, question would be to find the breaking point. There exist some papers on this topic using simulations, see e.g., [12, 2, 10] and further papers cited therein, but we are not aware of any theoretical results concerning this matter.

12.3 Remembering the first vs. 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 equals most likely the *first* one, that is, a typical path will then constist of an overwhelming amout of X_1:s interfoliated by an occasional $-X_1$. In the second case every new step equals most likely the *most recent* one, that is, a typical path will constist of an overwhelming amout of X_1:s followed by an overwhelming amount of $-X_1$:s, followed by ...., 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, etcetera, which implies that the next step, in fact, depends on the whole path so far.

12.4   Final Remarks

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

(ii) By using the device from Section 4 one can extend all limit theorems for ERW:s to the case with general steps.   □

A   Appendix

In this appendix we collect more technical calculations.

A.1   Proof of Lemma 8.1

Recall that even powers of X_n are always equal to 1, and, moreover, that $X_n^k = X_n$ if k is odd. One consequence of this is the following fact that will be used repeatedly below:

$$
\begin{aligned}
E\big((T_n)^{2m-k}(X_{n+1})^k \mid \mathcal{F}_n\big) &= E\big((T_n)^{2m-k}E(X_{n+1} \mid \mathcal{F}_n)\big) \\
&= \begin{cases} E(T_n)^{2m-k}, & \text{if } k \text{ is even,} \\ E\big((T_n)^{2m-k}(2p-1)X_n\big), & \text{if } k \text{ is odd.} \end{cases}
\end{aligned} \tag{A.1}
$$

As mentioned in connection with the statement of Theorem 8.1 we use induction. We thus assume that we know that the moments up to order $2m - 2$ converge properly, in particular we may choose n so large that $|E(T_n/\sqrt{n})^k| \le 2\mu_k$ for $k \le 2m - 2$, which, by symmetry, inplies that $|E(T_n)^k| \le 4\mu_k n^{k/2}$, for k even $\le 2m - 1$ and $\le 4\varepsilon n^{k/2}$, and for k odd $\le 2m - 1$, for some ε small (recall that μ_k are the moments of the standard normal distribution as given in (8.3)).
PROOF OF (8.4).

$$
\begin{aligned}
E\big((T_n)^{2m-1}X_n \mid \mathcal{F}_{n-1}\big) &= E\big((T_{n-1} + X_n)^{2m-1}X_n \mid \mathcal{F}_{n-1}\big) \\
&= \sum_{k=0}^{2m-1} \binom{2m-1}{k}(T_{n-1})^{2m-1-k} \cdot E\big((X_n)^{k+1} \mid \mathcal{F}_{n-1}\big) \\
&= (T_{n-1})^{2m-1}E(X_n \mid \mathcal{F}_{n-1}) + (2m-1)(T_{n-1})^{2m} \cdot E(X_n^2 \mid \mathcal{F}_{n-1}) + R_n(p,m) \\
&= (T_{n-1})^{2m-1}(2p-1)X_{n-1} + (2m-1)(T_{n-1})^{2m} + R_n(p,m).
\end{aligned}
$$

Taking expectations on either side yields

$$
E\big((T_n)^{2m-1}X_n\big) = (2p-1)E((T_{n-1})^{2m-1}X_{n-1}) + (2m-1)E(T_{n-1})^{2m-2} + E(R_n(p,m)). \tag{A.2}
$$

Exploiting (A.1) yields a bound for the remainder:

$$
\begin{aligned}
|E(R_n(p,m))| &\le \Big| \sum_{k=3}^{2m-1} \binom{2m-1}{k} E\big((T_{n-1})^{2m-1-k} \cdot E((X_n)^{k+1} \mid \mathcal{F}_{n-1})\big)\Big| \\
&\le \sum_{k=3}^{2m-1} \binom{2m-1}{k} |E(T_{n-1})^{2m-1-k}| \cdot 1 \\
&\le C_p \cdot 2m \cdot \binom{2m-1}{m-1} \cdot 4\mu_{2m-4} \cdot n^{m-2} = C(m,p) \cdot n^{m-2}. \tag{A.3}
\end{aligned}
$$

By iterating (A.2) we then obtain that

$$
\begin{aligned}
E\big((T_n)^{2m-1}X_n\big) &= (2m-1)\sum_{k=1}^{n-1}(2p-1)^{n-1-k}E\big((T_k)^{2m-2} \\
&\qquad + R_{k+1}(p,m)/(2m+1)\big) + (2p-1)^{n-1} \\
&= (2m-1)\frac{1}{2(1-p)}E(T_n)^{2m-2} + \mathcal{O}\big(n^{m-2}\big).
\end{aligned}
$$

PROOF OF (8.5).

$$E\big((T_n)^{2m}X_n \mid \mathcal{F}_{n-1}\big) = E\big((T_{n-1} + X_n)^{2m}X_n \mid \mathcal{F}_{n-1}\big)$$

$$= \sum_{k=0}^{2m}\binom{2m}{k}(T_{n-1})^{2m-k}\cdot E\big((X_n)^{k+1} \mid \mathcal{F}_{n-1}\big)$$

$$= (T_{n-1})^{2m}E(X_n \mid \mathcal{F}_{n-1}) + 2m(T_{n-1})^{2m-1}\cdot E(X_n^2 \mid \mathcal{F}_{n-1}) + R_n(p,m)$$

$$= (T_{n-1})^{2m}(2p-1)X_{n-1} + 2m(T_{n-1})^{2m-1} + R_n(p,m).$$

Taking expectations on either side yields

$$E\big((T_n)^{2m}X_n\big) = (2p-1)E\big((T_{n-1})^{2m}X_{n-1}\big) + 2mE(T_{n-1})^{2m-1} + E(R_n(p,m)).$$

The estimation of the remainder is the same as above. The remaining part of the proof follows the exact same lines and is therefore omitted.

Having estimates for the mixed moments we are now able to attack the "pure" moments. This will be done without explicit mentioning. Moreover, the estimates for the remainders, are, again, the same.

PROOF OF (8.6).

$$E\big((T_{n+1})^{2m} \mid \mathcal{F}_n\big) = \sum_{k=0}^{2m}\binom{2m}{k}(T_n)^{2m-k}\cdot E\big((X_{n+1})^k \mid \mathcal{F}_n\big)$$

$$= (T_n)^{2m} + 2m(2p-1)(T_n)^{2m-1}X_n + \binom{2m}{2}(T_n)^{2m-2} + R_n(p,m).$$

Taking expectations on either side yields

$$\begin{aligned}
E(T_{n+1})^{2m} &= E(T_n)^{2m} + 2m(2p-1)E\big((T_n)^{2m-1}X_n\big) + \binom{2m}{2}E(T_n)^{2m-2} + E(R_n(p,m))\\[2mm]
&= E(T_n)^{2m} + 2m(2p-1)\cdot(2m-1)\frac{1-(2p-1)^{n-1}}{2(1-p)}E(T_n)^{2m-2}\\[2mm]
&\quad + \binom{2m}{2}E(T_n)^{2m-2} + \mathcal{O}\big(n^{m-2}\big)\\[2mm]
&= E(T_n)^{2m} + \frac{m(2m-1)p}{1-p}\cdot E(T_n)^{2m-2} + \mathcal{O}\big(n^{m-1}\big)\\[2mm]
&= E(T_n)^{2m} + \frac{m(2m-1)p}{1-p}\frac{(2m-2)!}{2^{m-1}(m-1)!}n^{m-1} + \mathcal{O}\big(n^{m-2}\big),
\end{aligned}$$

by the induction hypothesis. Summing up the differences leads to the desired result.

PROOF OF (8.7).

$$E\big((T_{n+1})^{2m+1} \mid \mathcal{F}_n\big) = \sum_{k=0}^{2m+1}\binom{2m+1}{k}(T_n)^{2m+1-k}\cdot E\big((X_{n+1})^k \mid \mathcal{F}_n\big)$$

$$= (T_n)^{2m+1} + (2m+1)(2p-1)(T_n)^{2m}X_n + \binom{2m+1}{2}(T_n)^{2m-1} + R_n(p,m).$$

Taking expectations on either side yields

$$\begin{aligned}
E(T_{n+1})^{2m+1} &= E(T_n)^{2m+1} + (2m+1)(2p-1)E\big((T_n)^{2m}X_n\big)\\[2mm]
&\quad + \binom{2m+1}{2}E(T_n)^{2m-1} + E(R_n(p,m))\\[2mm]
&= E(T_n)^{2m+1} + (2m+1)(2p-1)C_1(p,m)\,n^{m-1}\\[2mm]
&\quad + m(2m-1)C_2(p,m-1)\,n^{m-1} + \mathcal{O}\big(n^{m-2}\big)
\end{aligned}$$

by the induction hypothesis. Summing up, finally, leads to the desired result with $C_2(p,m) = \frac{(2m+1)(2p-1)}{m}\cdot C_1(p,m) + (2m-1)\cdot C_2(p,m-1)$. $\qquad\square$

A.2 Proof of Lemma 9.1

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

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

For $n \geq 2$ we have

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

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

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

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

$$
\begin{aligned}
E(S_n) &= \sum_{k=1}^n \mu_k = \frac{a(3a + \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}} \frac{1 - \lambda_1^n}{1 - (a + \sqrt{a^2 + 4a})/2} \\
&\qquad - \frac{a(3a - \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}} \frac{1 - \lambda_2^n}{1 - (a - \sqrt{a^2 + 4a})/2} \\
&\to \frac{2a(3a + \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}(2 - a - \sqrt{a^2 + 4a})} - \frac{2a(3a - \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}(2 - a + \sqrt{a^2 + 4a})} \\
&= \frac{2a(a+1)}{1 - 2a} = \frac{p^2 - 1/4}{1 - p} \quad \text{as} \quad n \to \infty.
\end{aligned}
$$

The second moment is more tedious. We begin with

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

and obtain

$$v_{n+1} = E(S_{n+1}^2) = v_n + (2p-1)E(S_nX_n + S_nX_{n-1}) + 1. \tag{A.4}$$

As for the mixed moments,

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

By the usual trick we find

$$E(X_nX_{n-1}) = a + a^2 + \cdots + a^{n-1}E(X_2X_1) \to \frac{a}{1-a} = \frac{2p-1}{3-2p} \quad \text{as} \quad n \to \infty.$$

With $\zeta_n = E(S_nX_n)$ we find that

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

from which it follows that $\zeta_n \to \frac{p^2 - 2p + 7/4}{(1-p)(3-2p)}$, the stationary solution.

Next,

$$E(S_nX_{n-1}) = E(S_{n-1}X_{n-1}) + E(X_nX_{n-1}) = \zeta_{n-1} + \frac{a}{1-a} + \mathcal{O}(a^n).$$

We finally arrive, recalling (A.4), at

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

and thus, via telescoping, at

$$v_n \sim n\left(1 + \frac{(2p-1)(5-2p)}{2(1-p)(3-2p)}\right). \qquad \qquad \square$$

A.3   Calculation of second moments in Section 10

We first note that $E(T_1^2) = 1$, and, by (3.5), that, for $n \geq 1$,

$$E(T_{n+1}^2) = E(T_n^2) + (2p-1)E(T_n) + (2p-1)E(T_nX_n) + 1. \tag{A.5}$$

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

$$E(X_2X_1) = E\big(X_1E(X_2 \mid X_1)\big) = E(X_1(2p-1)X_1) = 2p-1,$$

so that

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

For $n \geq 2$ we exploit (3.4), (10.2), and the fact that $X_n^2 = 1$, to obtain

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

Another application of Proposition 3.2(i) then tells us that

$$
\begin{aligned}
E(T_nX_n) &= \Big\{ \frac{(2p-1)^2}{2(3-2p)} \cdot (n-1) + \frac{4(2p-1)(1-p)}{(3-2p)^2} + 1\Big) \Big/ \Big(1 - \frac{2p-1}{2}\Big) \\
&\qquad - \frac{2p-1}{2} \cdot \frac{(2p-1)^2}{2(3-2p)} \cdot (n-1) \Big/ \Big(\big(1 - \frac{2p-1}{2}\big)^2\Big\} \cdot \big(1 + o(1)\big) \\
&= \frac{(2p-1)^2}{(3-2p)^2} \cdot n + \frac{8(1+p-2p^2)}{(3-2p)^3} + o(1).
\end{aligned} \tag{A.6}
$$

Hence, using (A.5), we obtain

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

after which we, via telescoping, obtain that

$$
\begin{aligned}
E(T_n^2) &= \frac{(2p-1)^2}{(3-2p)^2} \cdot n^2 - \frac{(2p-1)^2}{(3-2p)^2} \cdot n + (n-1) \cdot \Big(1 + \frac{32(2p-1)(1-p)}{(3-2p)^3}\Big) + o(n). \\
&= \frac{(2p-1)^2}{(3-2p)^2} \cdot n^2 + \Big(1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 40p + 35)\Big) \cdot n + o(n).
\end{aligned} \tag{A.7}
$$

A.4   Calculation of second moments in Section 11

The point of departure in this case is (11.3), viz.,

$$E(T_{n+1}^2) = E(T_n^2) + 2E(T_nX_{n+1}) + 1. \tag{A.8}$$

For the mixed moments we use (3.4):

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

We thus find, using (11.2), that for $n \geq 3$,

$$
\begin{aligned}
E(T_nX_{n+1}) &= \frac{2p-1}{3}E(T_{n-1}X_n) + \frac{2p-1}{3} + \frac{2(2p-1)}{3}\Big(n \cdot \frac{2p-1}{2-p} + \frac{3(1-p)(7-2p)}{2(2-p)^2} + o(1)\Big) \\
&= \frac{2p-1}{3}E(T_{n-1}X_n) + \frac{2(2p-1)^2}{3(2-p)^2} \cdot n + \frac{2p-1)(1-p)(7-2p)}{(2-p)^2} + \frac{2p-1}{3} + o(1).
\end{aligned}
$$

Invoking Proposition 3.2(i) then tells us that

$$
\begin{aligned}
E(T_n X_{n+1}) &= \left\{ \frac{2(2p-1)^2}{3(2-p)^2} \cdot n + \frac{2p-1)(1-p)(7-2p)}{(2-p)^2} + \frac{2p-1}{3} \right) \Big/ \left(1 - \frac{2p-1}{3} \right) \right. \\
&\qquad \left. - \frac{2p-1}{3} \cdot \frac{2(2p-1)^2}{3(2-p)^2} \cdot n \Big/ \left((n+1)\Big(1 - \frac{2p-1}{3}\Big)^2 \right) \right\} \cdot \big(1 + o(1) \big) \\
&= \frac{(2p-1)^2}{(2-p)^2} \cdot n + \frac{3(2p-1)}{(2-p)^3} \cdot (p^2 - 9p + 8) + o(1) ,
\end{aligned}
$$

which, inserted into (A.8), yields

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

and, after summation,

$$
\begin{aligned}
E(T_n^2) &= n^2 \cdot \frac{(2p-1)^2}{(2-p)^2} - n \cdot \frac{(2p-1)^2}{(2-p)^2} + (n-1) \cdot \left(1 + \frac{3(2p-1)(p^2 - 9p + 8)}{(2-p)^3} \right) + o(n) \\
&= n^2 \cdot \frac{(2p-1)^2}{(2-p)^2} + n \cdot \left(1 + \frac{(2p-1)}{(2-p)^3} \cdot (5p^2 - 32p + 26) \right) + o(n). \qquad \text{(A.9)}
\end{aligned}
$$

We, finally, 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, $E(T_1 X_1) = 1$, $E(T_2 X_2) = E(X_1 X_2 + X_2^2) = -1 + 1 = 0$, and $E(T_3 X_3) = E(T_2 X_3 + X_3)^2 = 0 + 1 = 1$. For $n \geq 3$ we follow the usual pattern. Due to the fact that the mean is zero, an application of (3.4) now yields

$$
E(T_n X_{n+1}) = \frac{2p-1}{3} E(T_n X_n) = \frac{2p-1}{3} E(T_{n-1} X_n) + \frac{2p-1}{3}, \qquad \text{(A.10)}
$$

which, together with Proposition 3.2(i), tells us that

$$
E(T_n X_{n+1}) = \frac{\frac{2p-1}{3}}{1 - \frac{2p-1}{3}} + o(1) = \frac{2p-1}{2(2-p)} + o(1) \quad \text{as} \quad n \to \infty. \qquad \text{(A.11)}
$$

Moving into second moments, $E(T_1^2) = 1$, $E(T_2^2) = 0$, and $E(T_3^2) = E(X_3^2) = 1$. For $n \geq 3$ we insert our findings in (A.11) into (A.8):

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

so that, via telescoping,

$$
E(T_n^2) = n \cdot \frac{1+p}{2-p} + o(n) = \mathrm{Var}\,(T_{n+1}) \quad \text{as} \quad n \to \infty. \qquad \text{(A.12)}
$$

Acknowledgement

References

[1] Bercu, B. (2018). A martingale approach for the elephant random walk. *J. Phys. A: Math. Theor.* **81**, 015201.

[2] Cressoni, J.C., da Silva, M.A.A., and Viswanathan, G.M. (2007). Amnestically induced persistence in random walks *J. Phys. A.: Math. Theor.* **46**, 505002.

[3] DOOB, J.L.(1953). *Stochastic Processes.* J. Wiley & Sons, New-York.

[4] ENGLÄNDER, J., AND VOLKOV, S. (2018). Turning a coin over instead of tossing it. *J. Theor. Probab.* **31**, 1097-1118.

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

[6] HERKENRATH, U., IOSIFESCU, M., AND RUDOLPH, A. (2003) A note on invariance principles for iterated random functions. *J. Appl. Probab.* **40**, 834837.

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

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

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

[10] MOURA, TH.R.S., VISWANATHAN, G.M., DASILVA, M.A.A., CRESSONI, J.C., AND DASILVA, L.R. (2016). Transient superdiffusion in random walks with a *q*-exponentially decaying memory profile. *Physica A.: Statist. Mech. Appl.* **453**, 259-263.

[11] 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.

[12] DA SILVA, M.A.A., CRESSONI, J.C., AND VISWANATHAN, G.M. (2006). *Physica A* **364**, 70-78.

Allan Gut, Department of Mathematics, Uppsala University, Box 480, SE-751 06 Uppsala, Sweden;
Email: `allan.gut@math.uu.se`
URL: `http://www.math.uu.se/~allan`

Ulrich Stadtmüller, Ulm University, Department of Number Theory and Probability Theory,
D-89069 Ulm, Germany;
Email: `ulrich.stadtmueller@uni-ulm.de`
URL: `http://www.mathematik.uni-ulm.de/en/mawi/institute-of-number-theory-and-probability-theory/people/stadtmueller.html`