Journal of Mathematical Physics

Non-Markovian random walks with memory lapses

Cite as: J. Math. Phys. **59**, 113301 (2018); https://doi.org/10.1063/1.5033340
Submitted: 04 April 2018 . Accepted: 29 October 2018 . Published Online: 15 November 2018

Manuel González-Navarrete, and Rodrigo Lambert

JOURNAL OF MATHEMATICAL PHYSICS **59**, 113301 (2018)

Non-Markovian random walks with memory lapses

Manuel González-Navarrete[1,a)] and Rodrigo Lambert[2,b)]
[1]*Departamento de Estadística, Universidad del Bío-Bío, Concepción, Chile*
[2]*Faculdade de Matemática, Universidade Federal de Uberlândia and Departamento de Matemática, Universidade Federal da Bahia, Salvador, Brazil*

(Received 4 April 2018; accepted 29 October 2018; published online 15 November 2018)

We propose an approach to construct Bernoulli trials $\{X_i,\ i \geq 1\}$ combining dependence and independence periods, and we call it the Bernoulli sequence with random dependence (BSRD). The structure of dependence, in the past $S_i = X_1 + \cdots + X_i$, defines a class of non-Markovian random walks of recent interest in the literature. In this paper, the dependence is activated by an auxiliary collection of Bernoulli trials $\{Y_i,\ i \geq 1\}$, called *memory switch sequence*. We introduce the concept of *memory lapse property*, which is characterized by intervals of consecutive independent steps in BSRD. The main results include classical limit theorems for a class of linear BSRD. In particular, we obtain a central limit theorem for a class of BSRD which generalizes some previous results in the literature. Along the paper, several examples of potential applications are provided. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5033340

I. INTRODUCTION AND MOTIVATION

Independent and identically distributed Bernoulli trials $X_1, \ldots, X_n$ and their related random walks $S_n = X_1 + \cdots + X_n$ are among the most studied subjects in statistics and probability theories. The generalizations of such processes have also been widely investigated in the literature, either by removing the identically distributed hypothesis or by considering some dependence structure in the sequence.[8]

In this paper, we define a sequence $\{X_i,\ i \geq 1\}$ of Bernoulli random variables, in which each trial has probability of a success either as function of the number of previous successes S_n or independent of that. The dependence will be activated/inactivated by a latent collection of independent Bernoulli trials $\{Y_i,\ i \geq 1\}$, called the *memory switch sequence*. In other words, we construct Bernoulli sequences which have the flexibility to combine dependent and independent periods. The dependence will be in "on" (respectively, "off") mode, whenever the switch factor "$Y_i = 1$" (respectively, "$Y_i = 0$"). We call the model as Bernoulli sequences with random dependence (BSRD).

The dependence structure is taken from classical applications in epidemiological studies (see, for instance, Chap. 7 of Zelterman[24]). In addition, as noted in Ref. 1, a class of correlated random walks[3,4,13,19,21] (called *non-Markovian* in Ref. 21) can also be defined by that kind of dependence structure. In this sense, the Bernoulli random variable mean steps up or down if $X_i = 1$ or $X_i = 0$, respectively. Formally, let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$ be the σ-field generated by the sequence $X_1, \ldots, X_i$. The probability of a success on the $(i+1)$-th trial, given its past $\mathcal{F}_i$, satisfies $\mathbb{P}(X_{i+1} = 1|\mathcal{F}_i) = \mathbb{P}(X_{i+1} = 1|S_i)$. In words, the whole information of the past is summarized in S_i. In particular, for these processes, the present paper adopts the following notation:

$$P_i^*(s) = \mathbb{P}(X_{i+1} = 1|S_i = s) \tag{1}$$

a)magonzalez@ubiobio.cl
b)rodrigolambert@ufu.br

for $i \geq 1$ and $0 \leq s \leq i$. Then, Eq. (1) in view of the model proposed by Hod and Keshet[13] is given by

$$P_i^*(s) = \frac{1}{2}\left(1 - \mu\frac{i}{i+l}\right) + \mu\frac{s}{i+l},$$

where $-1 < \mu < 1$ and $l > 0$ is a *constant transient time* that obeys the following correlation rule: For $i \ll l$, the *past effect* is not too intense, while for $i \gg l$, it is. Another example is the elephant random walk (ERW) introduced in Ref. 21. In the ERW, it is supposed that the elephant remembers its full history and chooses its next step as follows. First, it selects a step uniformly from the past. Then, with probability $p \in [0, 1]$, it repeats what it did at the remembered time, whereas with the complementary probability $1 - p$, it makes a step in the opposite direction. In this respect, the probability of ith-step of ERW can be written as

$$P_i^*(s) = (1-p) + (2p-1)\frac{s}{i}, \tag{2}$$

where $p \in [0, 1]$. Moreover, the generalized binomial distribution proposed by Drezner and Farnum[6] is obtained in similar way by

$$P_i^*(s) = (1-\beta)p + \beta\frac{s}{i}, \tag{3}$$

where $0 \leq \beta \leq 1$ and $p = \mathbb{P}(X_1 = 1)$ is the initial probability.

We recall that in the present family of random walks, each step X_n is non-Markovian since it depends on all the previous steps. However, the position of the walk S_n evolves as a Markov chain. In fact, the conditional probabilities in (1) are related to some generalized Pólya-type urn[1,20] (for details, see Ref. 10). In this sense, the work of Janson[18] provides a general theory about convergence of this family of processes.

The present paper proposes a complementary approach to define the conditional probabilities of X_{i+1} given S_i, in view of (1). We present the BSRD and the property of *memory lapses* (see Definition 2), which is given by a string of 0's in $\{Y_i, i \geq 1\}$ that represents a period of independence in the steps of the BSRD $\{X_i, i \geq 1\}$. The main results are classical limit theorems for S_n in a class of BSRD with linear dependence. More specifically, we derive conditions on the parameters to obtain strong law of large numbers (SLLN), central limit theorem (CLT), and an invariance principle for S_n. We also generalize asymptotic results for some models in the literature, providing explicit limit quantities related to the asymptotic distributions.

The paper is organized as follows. In Sec. II, we define the BSRD and study some of its properties. Section III includes the main results. Section IV provides examples and discusses applications of the memory lapses. Finally, the tools and proofs of main theorems are given in Sec. V.

II. BERNOULLI SEQUENCES WITH RANDOM DEPENDENCE

Let $\{X_i, i \geq 1\}$ be a sequence of Bernoulli trials and $\{Y_i, i \geq 1\}$ an auxiliary collection of independent Bernoulli random variables, with $\mathbb{P}(Y_i = 1) = \lambda_i$ and independent of $S_i = X_1 + \cdots + X_i$, for all $i \geq 1$. The dependence structure of $\{X_i\}$ (for short notation) will be associated with the sequence $\{Y_i\}$, which we call the *memory switch sequence*. Formally, the probability of having a success at $(i+1)$-th trial is related to its past information S_i and the realization of the random variable Y_i by $\mathbb{P}(X_{i+1} = 1 | \mathcal{F}_i^X, \mathcal{F}_i^Y) = \mathbb{P}(X_{i+1} = 1 | S_i, Y_i)$, where $\mathcal{F}_i^X = \sigma(X_1, \ldots, X_i)$ is the σ-field generated by the sequence $\{X_i\}$, and similarly for $\mathcal{F}_i^Y$. Let us denote

$$P_i(s, y) = \mathbb{P}(X_{i+1} = 1 | S_i = s, Y_i = y), \tag{4}$$

for all $i \geq 1$, $0 \leq s \leq i$, and $y = 0, 1$. We recall that P is defined in a different probability space than P^* in (1). We remark that for $y = 0$, the probability $P_i(s, 0)$ will not depend on the previous successes S_i. However, if $y = 1$, the dependence exists. We think the random variable Y_i as a latent factor that determines the choice of dependence (or independence) on the past S_i for the trial X_{i+1}. Now, the definition for the so-called Bernoulli sequence with random dependence (BSRD) is provided.

Definition 1. A BSRD is the collection $\{X_i\}$, with memory switch sequence $\{Y_i\}$, defined by conditional probabilities (4) satisfying

$$P_i(s, 0) = P_i^*(0) \quad and \quad P_i(s, 1) = P_i^*(s), \tag{5}$$

where $P_i^*(\cdot)$, given by (1), is the probability of an embedded dependent Bernoulli sequence.

Let us now state a simple technical result that allows us to obtain finite-dimensional laws $\mathbb{P}(X_1, \ldots, X_n)$, for all $n \geq 1$, and then check the existence of the BSRD as a stochastic process. Consider a BSRD $\{X_i, 1 \leq i \leq n\}$, as in Definition 1; thus, using independence of Y_i and S_i combined with (4), we have

$$\begin{aligned} P_i(s) := \mathbb{P}(X_{i+1} = 1 | S_i = s) &= \sum_{y\in\{0,1\}} \mathbb{P}(X_{i+1} = 1 | S_i = s, Y_i = y)\mathbb{P}(Y_i = y) \\ &= \sum_{y\in\{0,1\}} P_i(s, y)\mathbb{P}(Y_i = y) \\ &= (1 - \lambda_i)P_i^*(0) + \lambda_i P_i^*(s). \end{aligned} \tag{6}$$

Now, let us take a look at BSRD as being a stochastic process. Particularly, we remark two important facts related to a realization of such a process [see (7) as reference].

- (R1) If $Y_i = 0$, then X_{i+1} is independent of its past. In other words, X_{i+1} is chosen from a Bernoulli distribution with parameter $\alpha_i = P_i^*(0)$ independent of the observation from $(X_1, \ldots, X_i)$.
- (R2) If $Y_i = 1$, then X_{i+1} should be dependent on its past. The probability $P_i^*(s)$ is defined by the embedded dependent Bernoulli sequence.

Note that by (R1), a period of independence in the BSRD is given by a string of 0's in the memory switch sequence. This fact leads us to the following definition:

Definition 2. A memory lapse in the BSRD $\{X_i\}$ is an interval $I \subset \mathbb{N}$ such that $Y_i = 0$ for all $i \in I$ and there is no interval $J \neq I$ with $J \supset I$ such that $Y_i = 0$ for all $i \in J$. The length of the lapse is $|I|$.

In some sense, we think this period as a lapse because after this, the model always will recover the dependence on the whole past, given by the conditional probability $P_i^*(s)$. More details about the notion of memory lapses, their applications, and some mathematical properties will be discussed in Sec. IV. We recall that the BSRD formulation generalizes the models studied in (1) in the sense that if $\lambda_i \equiv 1$ ($\lambda_i = 1$ for all $i \geq 1$), then there are no memory lapses and the evolution of the process is completely given by conditional probabilities (1). Instead, in the other extreme, if $\lambda_i \equiv 0$, the BSRD has the same probability measure as independent Bernoulli trials, with probability of success $P_i^*(0)$. In particular, if $P_i^*(0) = 1/2$, we obtain the well-known simple symmetric random walk.

III. MAIN RESULTS

We study, as embedded dependent processes, the family of linear-dependent Bernoulli sequences described in the work of Wu *et al.*[23] This leads us to an application of BSRD in a wide class of previous models in the literature. Using notation (4), we provide the following definition:

$$P_i(s, y) = \alpha_i + \beta_i \frac{sy}{i} \quad \text{and} \quad \mathbb{P}(X_1 = 1) = P_0(s) = \alpha_0, \tag{7}$$

for all $i \geq 1$, $0 \leq s \leq i$, and $y = 0, 1$. The sequences $\{\alpha_i\}$ and $\{\beta_i\}$ of parameters must satisfy the following conditions: $\alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i \leq 1$. We remark that β_i measures the strength of the dependence, while parameter $\lambda_i = \mathbb{P}(Y_i = 1)$ represents that the probability to that dependence actually exists.

In the following results, we show conditions to obtain classical limit theorems for S_n in the family of BSRD defined by (7). Let us first define some quantities which appear in the main results. We start with $a_1 = 1$ and for $n \geq 2$,

$$a_n = \prod_{k=1}^{n-1}\left(1 + \frac{\beta_k}{k}\lambda_k\right), \quad A_n^2 = \sum_{i=1}^{n} \frac{1}{a_i^2}, \quad \text{and} \quad B_n^2 = \sum_{i=1}^{n} \frac{p_i(1-p_i)}{a_i^2}, \tag{8}$$

where $p_i = \mathbb{P}(X_i = 1)$. Now we state the main theorems of this paper. The first one provides a strong law of large numbers (SLLN) for S_n.

Theorem 1. *Consider a BSRD $\{X_i\}_{1\le i\le n}$ as in (7). Then,*

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}(S_n)}{n} = 0 \quad a.s \tag{9}$$

if and only if

$$\sum_{k=1}^{\infty} \frac{1-\beta_k\lambda_k}{1+k} = \infty. \tag{10}$$

In (10), we recall that the relation between the real number sequences $\{\beta_i\}_{i\in\mathbb{N}}$ and $\{\lambda_i\}_{i\in\mathbb{N}}$ plays a central role in the SLLN. For instance, if $\beta_i = 1/\lambda_i$ for all i, we do not get the convergence (9).

In what follows, we present the second theorem. It provides necessary conditions on the quantities (8) in order to obtain an invariance principle for S_n.

Theorem 2. *Suppose in addition to the hypothesis of Theorem 1 that $\lim_{n\to\infty} B_n = \infty$ and $\limsup_{n\to\infty} A_n/B_n < \infty$. Then it is possible to redefine $\{X_i\}$ in a new probability space without changing its distribution, and there exists a standard Brownian motion $\{W(t)\}$ defined on the same probability space such that*

$$\text{(a)} \quad \frac{\left|\frac{S_n-\mathbb{E}(S_n)}{a_n} - W(B_n^2)\right|}{B_n\sqrt{\log\log B_n}} \xrightarrow{a.s.} 0 \quad , \quad \text{(b)} \quad \frac{\left|\frac{S_n-\mathbb{E}(S_n)}{a_n} - W(B_n^2)\right|}{B_n} \xrightarrow{\mathbb{P}} 0\,. \tag{11}$$

In (11), the symbols $\xrightarrow{a.s.}$ and $\xrightarrow{\mathbb{P}}$ mean almost-sure convergence and convergence in probability, respectively, with all limits taken as n diverges. We recall that the central limit theorem (CLT) and also the law of the iterated logarithm (LIL) for BSRD follow straightforwardly from Theorem 2 and the CLT and LIL for the standard Brownian motion.

We recall that, even in particular cases, the computation of marginal probabilities p_i in (8) is a non-trivial problem.[12] However, a relation between the BSRD in (4) and a class of Pólya-type urns allows us to generalize some asymptotic results for previous models in the literature, in particular, the generalized binomial studied in Refs. 6 and 17 as expressed in (3). The conditional probabilities (7) for the BSRD are given by

$$P_i(s, y) = (1-\beta)\alpha_0 + \beta y \frac{s}{i}, \tag{12}$$

where $0 \le \beta \le 1$ and $\alpha_0 = \mathbb{P}(X_1 = 1)$. Then, the following result provides explicit limiting proportion of successes for this particular class.

Theorem 3. *Let $\{X_i\}$ be a BSRD defined by conditional probabilities (12), with $\{Y_i\}$ independent and identically distributed (i.i.d.) parameter $\lambda \in [0.1]$. Then*

(i) *If $\beta\lambda < 1/2$,*

$$\frac{1}{\sqrt{n}}\left(S_n - \frac{n\alpha_0(1-\beta)}{1-\beta\lambda}\right) \xrightarrow{d} N(0,\sigma^2) \quad \text{as } n\to\infty\,, \tag{13}$$

where

$$\sigma^2 = \frac{\alpha_0(1-\alpha_0-\beta(\lambda-\alpha_0))(1-\beta)}{(1-2\beta\lambda)(1-\beta\lambda)^2}. \tag{14}$$

(ii) *If $\beta\lambda = 1/2$,*

$$\frac{S_n - 2n\alpha_0(1-\beta)}{\sqrt{n\log n}} \xrightarrow{d} N(0,\sigma^2) \quad \text{as } n\to\infty\,, \tag{15}$$

where

$$\sigma^2 = 4\alpha_0\left(\frac{1}{2} - \alpha_0(1-\beta)\right)(1-\beta)\,. \tag{16}$$

The proofs of all theorems will be given in Sec. V and are based on the classical results for convergence of bounded martingale differences (Theorems 1 and 2) and the relation of BSRD with a class of generalized Pólya urns (Theorem 3).

IV. THE MEMORY LAPSES PROPERTY

This section presents examples and discusses potential applications for the BSRD. The notion of memory lapses is exploited in the context of each example/application. The goal is to discuss the flexibility provided by this property in further mathematical modeling of real problems.

Consider a contagious disease affecting a finite population. Each of the individuals can be described by Bernoulli distributed indicators of their disease status. In other words, $X_i = 1$ if the ith member is a diseased case and X_i is zero if this member is otherwise healthy. The collection of Bernoulli trials $\{X_i, i \geq 1\}$ has been extensively studied in the literature by conditional probabilities (1). We refer to, for instance, Ref. 24.

In the approach of BSRD, we suppose that a genetic or environmental factor can be represented by the memory switch sequence $\{Y_i\}$. That is, such factor activates/inactivates the dependence structure of the model. Then, for instance, consider the model in (12) and the following two situations.

Example 1. Let us denote the switch factor $Y_i = \mathbb{I}_{\{Z_i<z\}}$*, where* Z_i *is a continuous random variable which represents the concentration of an "immunizing antibody" related to the (*i + 1*)-th individual and* z *is the critical value for immunization. Then, if* $Z_i \geq z$*, the probability of the (*i + 1*)-th member becoming diseased is independent of the history of such a disease in the population. Otherwise, the probability will increase by the (familiar) history of successes* S_i*, the number of sick individuals until time* i*.*

Example 2. In other situations, Y_i *may denote a vaccine to prevent infection, where* $Y_i = 0$ *means that the vaccine is effective when applied on the (*i + 1*)-th member of the population. Otherwise, the vaccine does not actuate and the probability increases by the history of sick individuals.*

Therefore, as in Definition 2, a memory lapse of length l could be interpreted as a group of l consecutive patients for whom the vaccine is effective. That is, the history of sickness is neglected.

In what follows, we give other examples trying to understand better the memory lapse property in different situations.

Example 3. Customers can buy a product ($X_i = 1$*) by necessity or due to another reason. Thinking on it, an advertising program* $\{Y_i\}$ *is launched that can or cannot reach the customers. Consider the conditional probabilities (7). Let be* α_i *the quantity related to necessity and* β_i *the quantity related to the "social pressure" (history of selling) of the product as defined in Ref. 2, associated with the (*i + 1*)-th customer. Then, if* $Y_i = 1$*, that is, the publicity reached him/her, the "social pressure" will raise the probability of the customer buying the product [see remark (R2) in Sec. II].*

Example 4. Imagine that there is a robot (computer) used to buy/sell some asset in the stock market. It obeys an algorithm that determines such decision. At the (i + 1*)-th decision, the robot tries to access the history of buying/selling activity. If it has success (*$Y_i = 1$*), then its decision will be based on the history of the asset. If it does not, due to some noise or stoppage, for instance, then it decides to buy (or sell) the asset only by flipping a coin. Here we are supposing that although its memory is inaccessible (*$Y_i = 0$*), it has to take a decision anyway.*

Example 5. (The hot hand in basketball) In basketball, there is a common belief that the probability of hitting a shot after a hit is greater than the probability of hitting after a miss. Moreover, a sequence of consecutive hits will increase the probability of a hit in the next shot. This is the so-called "hot hand phenomenon" (we refer the reader to the paper by Gilovich et al.[9]). In the context of BSRD, the hot hand can be regarded as a consecutive string of 1's in $\{X_i\}$ *and* $\{Y_i\}$ *(or a string without misses and without memory lapses). Then the probability of hitting will rise since the basketball player comes from a success history.*

Therefore, in Example 3, a memory lapse is denoted by consecutive costumers for whom the publicity does not reach. In a similar sense, the memory lapses can be included in the model on social behavior about soft technologies introduced by Bendor *et al.*[2] In the case of Example 4, the memory lapses are periods of decisions taken without looking at the history of selling of the asset, described by interruptions in the system or the intranet connection. In view of Example 5, the memory lapse can be thought as a period in which the opposite team moves the defence to the player, which is a common strategy when the objective is to stop some player in hot hand.

Here we point out that the paths covered by the latent sequence $\{Y_i\}_{i=1}^n$ determines how many times the history of the process was taken into account in the construction of the path $\{X_i\}_{i=1}^n$. Section IV A presents a more detailed analysis of the switch sequence and provides some examples that illustrate its influence on the BSRD model.

A. Characterization of switch sequence

In a complementary line, we obtain some mathematical results about the occurrence of memory lapses in a BSRD. In particular, we analyze the property of memory lapses by studying random variables defined on $\{Y_i\}$, which we call the *switch sequence*.

First of all, note that in Secs. II and III, we supposed the trials $\{Y_i\}$ being independent, with $\mathbb{P}(Y_i = 1) = \lambda_i$. We highlight here that the problem to study patterns in an independent Bernoulli sequence (Poisson trials) has been widely studied in the literature.[8] Therefore there are several possibilities to analyze such patterns by assuming different forms for the parameter collection $\{\lambda_i\}$. The assumptions about those parameters are given by the specific application of the BSRD. However, as an illustration, we focus on two particular cases and compare their behaviors. The results are obtained by using techniques from recent studies in the study of pattern strings in Bernoulli sequences (see, for instance, Refs. 14–16 and 22).

In the first part, we study waiting times for the switch sequence $\{Y_i\}$ based on the approach introduced in Ref. 7. In this sense, we focus on the so called *frequency (FQ) and succession (SQ) quotas* for Bernoulli trials. Revisiting the situations given above, we think of applications to the analysis of these FQ and SQ problems.

Formally, in the case of FQ, we denote W_r^{FQ} as the waiting time until r failures (or 0's) have been observed in $\{Y_i\}$. We assume the memory switch sequence being i.i.d. In this case, we count the number of independent Bernoulli trials until observe r failures. Then, the random variable W_r^{FQ} follows a negative binomial distribution with parameters $1 - \lambda$ and r.

Example 6. (Relation between FQ and Example 3) Suppose that there is a criterion to evaluate publicity. For instance, if it completes a fixed quantity r of non-reached customers, then the publicity is removed. That is, we are interested in the waiting time until we attain a given quota of 0's. As we said above, the decision to remove publicity should be regarded as the probability distribution of a negative binomial random variable.

Note also that, if we want to see the proportion of failures, then we will deal with a binomial distribution. To illustrate this, suppose that we choose a "critical proportion" of failures $\delta \in (0, 1)$. Then the probability of reaching this proportion will be given by $P(Z_n \le \lfloor n\delta \rfloor)$, where $Z_n \sim Bin(n, 1 - \lambda)$. Here we recall that by $\lfloor x \rfloor$, we mean the greatest integer less or equal than x.

On the other hand, we consider a succession quota (SQ) problem. Let W_s^{SQ} be the waiting time until the first memory lapse of size s is observed. It is possible to obtain a probability generating function for W_s^{SQ} given by

$$\phi_W(t) = \frac{(1 - qt)(qt)^s}{(1 - \lambda t)(1 - qt) - \lambda q t^2 (1 - (qt)^{s-1})}, \tag{17}$$

where $q = 1 - \lambda$ (see Ref. 7). In particular, we obtain $\mathbb{E}(W_s^{SQ}) = (1 - q^s)/(\lambda q^s)$.

However, as other situations from SQ, in the next example, we do not look at the waiting time of a memory lapse. Instead, we analyze a pattern of 1's, which gives us particular information about the situation under study. Of course, if one interchanges λ by q in (17), we obtain the corresponding probability generating function.

Example 7. (Relation between SQ and Example 2) We define the waiting time given by the first instant that we obtain s consecutive 1's in the memory switch sequence. In other words, we are interested in the first time such that the vaccine has no effects in s consecutive patients.

Let us now address a second kind of question. We modify the approach to study pattern behaviors of the *memory lapses*. Then, define the random variables $M_l(n)$, $l \in \{1, 2, \ldots, n\}$, representing the number of memory lapses of length l in the first n trials of $\{Y_i\}$. Formally,

$$\begin{aligned} M_l(n) = &\sum_{k=1}^{n-l-1} Y_k(1 - Y_{k+1}) \cdots (1 - Y_{k+l})Y_{k+l+1} \\ &+ (1 - Y_1) \cdots (1 - Y_l)Y_{l+1} \ + \ Y_{n-l}(1 - Y_{n-l+1}) \cdots (1 - Y_n), \end{aligned} \tag{18}$$

for $n > l$, where the second and third terms represent a memory lapse at the beginning and at the end of the sequence, respectively. In other words, $M_l(n)$ is the number of runs of 0's (see Refs. 14 and 16) of length l in the first n trials of $\{Y_i\}$.

If we are able to obtain information about random variables given in (18), we could say something about, for instance, the groups of consecutive customers for whom the publicity does not reach in Example 3. Also with $M_l(n)$, we count the strings of l consecutive decisions without looking for the data in Example 4.

In addition to (18), we will be interested in periods of alternate dependence of BSRD. Formally, as functions of sequence $\{Y_i\}$, we have

$$A_l(n) = \sum_{k=1}^{n-2l+1} Y_k(1 - Y_{k+1}) \cdots Y_{k+2l-2}(1 - Y_{k+2l-1}) \tag{19}$$

being the number of alternating dependence-independence periods of size $2l$ in the first n trials of the BSRD $\{X_i\}$, where $n > 2l$. These random variables could give us information about periods of high interference in Example 4. Of course their characterization complements the information given by (18).

In the next result, we obtain the expectation of random variables $M_l(n)$ and $A_l(n)$. We consider the assumptions i.i.d. and the collection $\lambda_i = a/(a + b + i - 1)$ for $a > 0$, $b \geq 0$, and $i \geq 1$, usually denoted by $Bern(a, b)$ (see Refs. 15 and 22).

Proposition 1. (i) *If $\{Y_n\}$ are i.i.d. with parameter λ, then*

$$\mathbb{E}(M_l(n)) = (1 - \lambda)^l[2\lambda + \lambda^2(n - l - 1)], \quad l < n, \tag{20}$$

and

$$\mathbb{E}(A_l(n)) = (n - 2l)(\lambda(1 - \lambda))^l, \quad 2l < n. \tag{21}$$

(ii) *If $\{Y_n\}$ are Bern$(1, b)$, then*

$$\mathbb{E}(M_l(n)) = \frac{2b + l}{(b + l)(b + l + 1)}, \quad 1 \leq l < n, \tag{22}$$

and

$$\mathbb{E}(A_l(n)) = \sum_{k=1}^{n-2l+1} \prod_{i=1}^{l} \frac{1}{k + b + 2i - 1} \quad 2l < n. \tag{23}$$

Proof. Note that by independence, we have from Eq. (18),

$$\begin{aligned} \mathbb{E}(M_l(n)) = &\sum_{k=1}^{n-l-1} \lambda_k(1 - \lambda_{k+1}) \cdots (1 - \lambda_{k+l})\lambda_{k+l+1} \\ &+ (1 - \lambda_1) \cdots (1 - \lambda_l)\lambda_{l+1} + \ \lambda_{n-l}(1 - \lambda_{n-l+1}) \cdots (1 - \lambda_n). \end{aligned} \tag{24}$$

In part (i), the cases $\{Y_n\}$ are i.i.d., and by a straightforward calculation, we obtain (20). Similarly, we prove (21). Now, in part (ii), if $\{Y_n\}$ are $Bern(1, b)$, we should calculate

$$\mathbb{E}(M_l(n)) = \sum_{k=1}^{n-l-1}\left(\frac{1}{k+b+l} - \frac{1}{k+b+l+1}\right) + \frac{b}{(b+l)(b+l+1)} + \frac{1}{n+b}. \tag{25}$$

Then, solving the telescoping sum, we obtain (22). By similar arguments, we get $\mathbb{E}(A_l(n))$. □

We remark that it is possible to obtain second moments as recursive functions of expectations in Proposition 1. Note that (22) does not depend on the value of n; this is an interesting feature of Poisson trials $Bern(a, b)$, which have been studied in different contexts. For instance, the sequence $Bern(1, 0)$ arises in the limit in the study of cycles in random permutations and records values of continuous random variables. Moreover, the sequence $Bern(a, 0)$ has some applications in nonparametric Bayesian inference and species allocation models (see Refs. 15 and 22 and references therein).

V. PROOF OF MAIN RESULTS

The Proofs of Theorems 1 and 2 are based on the general results about convergence of martingale differences (see Ref. 11). First, let

$$M_n = \frac{S_n - \mathbb{E}(S_n)}{a_n}, \tag{26}$$

where a_n are given by (8), and denote

$$D_1 = M_1 \qquad D_n = M_n - M_{n-1}, \quad n \geq 2. \tag{27}$$

We aim to prove that $\{D_n, \mathcal{F}_n, n \geq 1\}$ is a sequence of bounded martingale differences. First, we need the particular cases of Theorem 2.17 and Corollary 3.1 included in the work of Hall and Heyde.[11] We state the results without proof.

Lemma 1. Let $\{Z_n, \mathcal{F}_n, n \geq 1\}$ be a sequence of martingale differences. If $\sum_{n=1}^{\infty} \mathbb{E}[Z_n^2|\mathcal{F}_{n-1}] < \infty$ a.s., then $\sum_{i=1}^{n} Z_i$ converges almost surely.

In order to check conditions in the previous lemma, we need the next auxiliary result, which is stated as follows:

Lemma 2. $\{D_n, \mathcal{F}_n, n \geq 1\}$ in (27) are bounded martingale differences.

Proof. First, we prove that M_n in (26) is a martingale,

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \frac{1}{a_{n+1}}\mathbb{E}[S_n + X_{n+1} - \mathbb{E}(S_n) - \mathbb{E}(X_{n+1})|\mathcal{F}_n] \\ &= \frac{1}{a_{n+1}}(S_n + \mathbb{E}[X_{n+1}|\mathcal{F}_n] - \mathbb{E}(S_n) - \mathbb{E}(X_{n+1})),\end{aligned} \tag{28}$$

by noticing that

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \sum_{y\in\{0,1\}} \mathbb{P}(X_{n+1} = 1|\mathcal{F}_n, Y_n = y)\mathbb{P}(Y_n = y) \\ &= (1-\lambda_n)\alpha_n + \lambda_n\left(\alpha_n + \frac{\beta_n}{n}S_n\right) \\ &= \alpha_n + \frac{\beta_n}{n}\lambda_n S_n\end{aligned} \tag{29}$$

and, given the independence of Y_n and S_n,

$$\begin{aligned}\mathbb{E}(X_{n+1}) &= \sum_{y\in\{0,1\}}\sum_{s=0}^{n} \mathbb{P}(X_{n+1} = 1|S_n = s, Y_n = y)\mathbb{P}(S_n = s)\mathbb{P}(Y_n = y) \\ &= (1-\lambda_n)\alpha_n + \lambda_n(\alpha_n\mathbb{P}(S_n = 0) + \cdots + (\alpha_n + \beta_n)\mathbb{P}(S_n = n)) \\ &= \alpha_n + \frac{\beta_n}{n}\lambda_n\sum_{i=0}^{n} i\mathbb{P}(S_n = i) = \alpha_n + \frac{\beta_n}{n}\lambda_n\mathbb{E}(S_n).\end{aligned} \tag{30}$$

Then $\{M_n\}_{n\in\mathbb{N}}$ is a martingale. Now, observe that for all $n \geq 2$,

$$D_n = \frac{X_n - \mathbb{E}(X_n)}{a_n} - \frac{S_{n-1} - \mathbb{E}(S_{n-1})}{n-1}\frac{\beta_{n-1}}{a_n}\lambda_{n-1} \tag{31}$$

and

$$|D_n| \leq \frac{2}{a_n}, \quad n \geq 1. \tag{32}$$

In other words, $\{D_n, \mathcal{F}_n, n \geq 1\}$ are bounded martingale differences. □

Proof of Theorem 1. First of all, note that

$$\frac{a_n}{n} = \frac{1}{n}\prod_{k=1}^{n-1}\left(\frac{k+\beta_k\lambda_k}{k}\right) = \prod_{k=1}^{n-1}\left(\frac{k+\beta_k\lambda_k}{k+1}\right) \tag{33}$$

since $\beta_k\lambda_k \leq 1$; thus, a_n/n is non-increasing. Moreover, $\frac{a_n}{n} = \exp(\sum_{k=1}^{n-1}\log(\frac{k+\beta_k\lambda_k}{k+1})) = \exp(-\sum_{k=1}^{n-1}\frac{1-\beta_k\lambda_k}{k+1} + O(1))$. Then, $\lim_{n\to\infty}a_n/n = 0$ if and only if $\sum_{k=1}^{\infty}\frac{1-\beta_k\lambda_k}{k+1} = \infty$.

Initially assume (10). That implies that $\frac{a_n}{n} \to 0$. Given Lemma 2, define the martingale differences $Z_i = \frac{a_i}{i}D_i$. By (32) $\sum_{i=1}^{\infty}\mathbb{E}[Z_i^2|\mathcal{F}_{i-1}] \leq \sum_{i=1}^{\infty}\frac{4}{i^2} < \infty$ a.s., Lemma 1 implies that $\sum_{i=1}^{\infty}Z_i$ converges a.s. Now we use Kronecker's lemma to obtain (9), that is,

$$\frac{a_n}{n}M_n = \frac{a_n}{n}\sum_{i=1}^{n}D_i \to 0 \text{ a.s. as } n \to \infty. \tag{34}$$

On the other hand, suppose $\sum_{k=1}^{\infty}\frac{1-\beta_k\lambda_k}{1+k} < \infty$. Applying (33) to, say, $v = \lim_{n\to\infty}\frac{a_n}{n} \in (0.1]$ and using (32), $\sum_{i=1}^{\infty}\mathbb{E}[D_i^2|\mathcal{F}_{i-1}] \leq \sum_{i=1}^{\infty}\frac{4}{a_i^2} < \infty$ a.s. Again, from Lemma 1, we obtain $M_n = \sum_{i=1}^{\infty}D_i$ converges a.s. to some random variable M, with $Var(M) = \lim_{n\to\infty}Var(M_n) = \sum_{i=1}^{\infty}\mathbb{E}(D_i^2) > 0$. Hence, M is a non-degenerate random variable. Thus, $\frac{a_n}{n}M_n$ also converges to a non-degenerate random variable and this completes the proof. □

Proof of Theorem 2. This proof will mainly follow the scheme presented in Ref. 25. Proving item (a), the Skorohod embedding theorem allows us to redefine $\{X_n, \mathcal{F}_n\}$ in a new probability space such that there is a Brownian motion $\{W(t)\}$ and a $\mathcal{F}_n-$filtered sequence of random variables $\tau_n \geq 0$ such that $M_n \overset{d}{=} W(T_n)$ (we can assume that $M_n = W(T_n)$ without loss of generality), where $T_n = \sum_{i=1}^{n}\tau_i$. Furthermore,

$$\mathbb{E}[\tau_j|\mathcal{F}_{j-1}] = \mathbb{E}[D_j^2|\mathcal{F}_{j-1}] \quad \text{and also} \quad \mathbb{E}[\tau_j^p|\mathcal{F}_{j-1}] = \mathbb{E}[D_j^{2p}|\mathcal{F}_{j-1}] \text{ a.s.}$$

Since $n^{-1}(S_n - \mathbb{E}(S_n)) \xrightarrow{a.s.} 0$, we get that

$$\mathbb{E}[D_j^2|\mathcal{F}_{j-1}] = \frac{p_j(1-p_j)}{a_j^2} + o_{a.s.}\left(\frac{1}{a_j^2}\right),$$

which implies that

$$\sum_{j=1}^{n}\mathbb{E}[\tau_j|\mathcal{F}_{j-1}] = B_n^2 + o_{a.s.}(A_n^2) = B_n^2 + o_{a.s.}(B_n^2)\ .$$

Here we recall that by "y_n is $o_{a.s.}(x_n)$," we mean $x_n/y_n \xrightarrow{a.s.} 0$. By noticing that $\hat{\tau}_j :\overset{d}{=} \tau_j - \mathbb{E}(\tau_j|\mathcal{F}_{j-1})$ is a sequence of martingale differences with respect to $\mathcal{F}_j$ and since $|D_j| \leq 2/a_j$, we conclude that $\mathbb{E}(\hat{\tau}_j|\mathcal{F}_{j-1}) \leq Ca_j^{-4}$, for some positive constant C. This implies that $A_j^{-4}\mathbb{E}(\hat{\tau}_j^2|\mathcal{F}_{j-1})$ is summable.

Now we combine Theorem 2.18 from Ref. 11 with Kronecker's lemma to conclude that $\sum_{j=1}^{n}\hat{\tau}_j = o_{a.s.}(A_n^2) = o_{a.s.}(B_n^2)$, which in turn implies that $T_n = B_n^2 + o_{a.s.}(B_n^2)$. Then we apply Theorem 1.2.1 of Ref. 5 combined with the fact that B_n diverges to obtain

$$M_n = W(T_n) = W(T_n^2) + o_{a.s.}(B_n\sqrt{\log\log B_n})\ .$$

Finally we divide both sides of the last equality above by $B_n\sqrt{\log\log B_n}$ and let n go to infinity to conclude the proof.

For item (b), the goal is to prove that for any $\epsilon > 0$, the probability $\mathbb{P}\left(\frac{|W(T_n)-W(B_n^2)|}{B_n} > \epsilon\right)$ vanishes as n diverges. For any $\delta \in (0, 1)$, we can decompose the above probability to get $\mathbb{P}\left(\frac{|W(T_n)-W(B_n^2)|}{B_n} > \epsilon, \frac{|T_n-B_n^2|}{B_n^2} > \delta\right) + \mathbb{P}\left(\frac{|W(T_n)-W(B_n^2)|}{B_n} > \epsilon, \frac{|T_n-B_n^2|}{B_n^2} \leq \delta\right)$.

It is straightforward to see that the first term above vanishes as $n \to \infty$. The second one can be bounded above by $\mathbb{P}(\limsup_{|s-1|\leq\delta} |W(s) - W(1)| > \epsilon)$, which also goes to zero as $\delta \to 0$ by the Lévy modulus of continuity for Wiener processes. For a small enough δ, let $n \to \infty$ to conclude the proof. □

Proof of Theorem 3. The strategy for this proof is to link the BSRD model to a generalized Pólya urn problem and then use the results stated by Janson in Ref. 18.

The first step consists in relating the distribution of $\{X_n\}_{n\in\mathbb{N}}$ with the distribution of the red balls in a two-color Pólya urn (namely, $\{R_n\}_{n\in\mathbb{N}}$). First of all, let us construct the *random replacement matrix* for the generalized Pólya urn. Here we will follow the notation given in Ref. 18. For this, consider the two column replacement vectors $\xi_1 = (\xi_{11}, \xi_{12})$ (red) and $\xi_2 = (\xi_{11}, \xi_{12})$ (blue), with $\xi_i \in \{(0, 1), (1, 0)\}$ (a single ball is replaced at each time), and the random replacement matrix given by $M = (\xi_1; \xi_2)$. Then if we choose the replacement vector ξ_1 to reinforce the urn, it means that we will replace ξ_{11} red balls and ξ_{12} blue balls. Otherwise we choose vector ξ_2 and then replace ξ_{21} red and ξ_{22} blue balls.

At each step, a ball is drawn from the urn, its color is observed, and the ball is replaced. The replacement column vector is then chosen according to the color of the withdrawn ball.

Note that $\mathbb{P}(\xi_{ij} = 1) = \mathbb{E}(\xi_{ij})$, for all i, j. If at time n we get r red balls ($R_n = r$), then the probabilities of replacing a red ball (success) or a blue ball (failure) at time $n + 1$ (conditioned on a proportion r/T_n of red balls) are, respectively, given by

$$\begin{aligned}\mathbb{P}(R_{n+1} = r + 1|R_n = r) &= \mathbb{P}(\xi_{11} = 1)\frac{r}{T_n} + \mathbb{P}(\xi_{21} = 1)\left(1 - \frac{r}{T_n}\right)\\ &= \mathbb{E}[\xi_{21}] + (\mathbb{E}[\xi_{11}] - \mathbb{E}[\xi_{21}])\frac{r}{T_n}\end{aligned} \tag{35}$$

and

$$\mathbb{P}(R_{n+1} = r|R_n = r) = \mathbb{E}[\xi_{22}] + (\mathbb{E}[\xi_{12}] - \mathbb{E}[\xi_{22}])\frac{r}{T_n}\ , \tag{36}$$

where $T_n = R_n + B_n$ is the total number of balls at time n (with B_n being the number of blue balls). Now we obtain the transition probabilities for $\{X_n\}$. By the independence of Y_n and S_n, we can repeat the argument used in the Proof of Lemma 2 (sum in all sets $\{Y_n = y\}$) to obtain from (12) the following probabilities (conditioned on a proportion s/n of previous successes):

$$\mathbb{P}(X_{n+1} = 1|S_n = s) = \alpha_0 - \alpha_0\beta + \lambda\beta\frac{s}{n}, \tag{37}$$

$$\mathbb{P}(X_{n+1} = 0|S_n = s) = 1 - \alpha_0 + \alpha_0\beta - \lambda\beta\frac{s}{n}. \tag{38}$$

The next step is to construct a matrix A given by

$$A = \mathbb{E}(M) = \begin{pmatrix}\mathbb{E}(\xi_{11}) & \mathbb{E}(\xi_{21})\\ \mathbb{E}(\xi_{12}) & \mathbb{E}(\xi_{22})\end{pmatrix}.$$

Then we relate (35) with (37) and (36) with (38) to obtain the expected replacement matrix

$$A = \begin{pmatrix}\alpha_0 + \beta(\lambda - \alpha_0) & \alpha_0 - \beta\alpha_0\\ 1 - \alpha_0 - \beta(\lambda - \alpha_0) & 1 - \alpha_0 + \beta\alpha_0\end{pmatrix}.$$

Now we obtain the key quantities to the limiting theorems stated in Ref. 18. As is known by that paper, the limiting theorems depend on the eigendecomposition of A. The two eigenvalues of A are $\ell_1 = 1$ and $\ell_2 = \beta\lambda$, and $v_1 = \frac{1}{1-\beta\lambda}\begin{pmatrix}\alpha_0(1-\beta)\\ 1-\alpha_0-\beta(\lambda-\alpha_0)\end{pmatrix}$ is the eigenvector associated with ℓ_1.

Now all the calculations are done by using the results stated in Ref. 18 (and examples therein). For item (i), we apply Theorem 3.22 of Ref. 18 combined with Lemmas 5.4 and 5.3 (i) of the same paper to obtain (13) and (14). For item (ii), we apply Theorem 3.23 of the same paper to obtain (15) and (16).

These last two arguments conclude the proof. □

ACKNOWLEDGMENTS

The authors thank G. Ludwig, T. Vargas, M. Abadi, and G. Ost for several comments on the original version (arXiv:1607.08299v1). M.G.N. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (Grant No. 2015/02801-6). He thanks kind hospitality and financial support from FAMAT-UFU. R.L. is partially supported by FAPESP (Grant No. 2014/19805-1) and CNPQ PDJ Grant (Process No. 406324/2017-4). He thanks the Center for Neuromathematics (IME-USP) for warm hospitality. They also thank the anonymous referee for helpful comments and corrections.

[1] Baur, E. and Bertoin, J., "Elephant random walks and their connection to Pólya-type urns," Phys. Rev. E **94**, 052134 (2016).
[2] Bendor, J., Huberman, B., and Wu, F., "Management fads, pedagogies, and other soft technologies," J. Econ. Behav. Organ. **72**, 290–304 (2009).
[3] Bercu, B., "A martingale approach for the elephant random walk," J. Phys. A: Math. Theor. **51**(1), 015201 (2018).
[4] Colleti, C., Gava, R., and Schütz, G., "Central limit theorem for the elephant random walk," J. Math. Phys. **58**, 053303 (2017).
[5] Csörgö, M. and Révész, P., *Strong Approximations in Probability and Statistics* (Academic Press, New York, 1981).
[6] Drezner, Z. and Farnum, N., "A generalized binomial distribution," Commun. Stat. Theory Methods **22**, 3051–3063 (1993).
[7] Ebneshahrashoob, M. and Sobel, M., "Sooner and later waiting time problems for Bernoulli trials: Frequency and run quotas," Stat. Probab. Lett. **9**, 5–11 (1990).
[8] Feller, W., *An Introduction to Probability Theory and Its Applications* (John Wiley, New York, 1968), Vol. 1.
[9] Gilovich, T., Vallone, R., and Tversky, A., "The hot hand in basketball: On the misperception of random sequences," Cognit. Psychol. **17**(3), 295–314 (1985).
[10] González-Navarrete, M. and Lambert, R., "Urn models with two types of strategies," preprint arXiv:1708.06430 (2017).
[11] Hall, P. and Heyde, C. C., *Martingale Limit Theory and Its Application* (Academic Press, New York, 1980).
[12] Heyde, C. C., "Asymptotics and criticality for a correlated Bernoulli process," Aust. N. Z. J. Statist. **46**, 53–57 (2004).
[13] Hod, S. and Keshet, U., "Phase transition in random walks with long-range correlations," Phys. Rev. E **70**, 015104(R) (2004).
[14] Holst, L., "On consecutive records in certain Bernoulli sequences," J. Appl. Probab. **46**, 1201–1208 (2009).
[15] Holst, L., "A note on records in a random sequence," Ark. Mat. **49**, 351–356 (2011).
[16] Holst, L. and Konstantopoulos, T., "Runs in coin tossing: A general approach for deriving distributions for functionals," J. Appl. Probab. **52**, 752–770 (2015).
[17] James, B., James, K., and Qi, Y., "Limit theorems for correlated Bernoulli random variables," Stat. Probab. Lett. **78**(15), 2339–2345 (2008).
[18] Janson, S., "Functional limit theorems for multitype branching processes and generalized Pólya urns," Stoch. Process. Appl. **110**, 177–245 (2004).
[19] Kürsten, R., "Random recursive trees and the elephant random walk," Phys. Rev. E **93**, 032111 (2016).
[20] Mahmoud, H., *Pólya Urn Models* (CRC Press, Boca Raton, 2008).
[21] Schütz, G. and Trimper, S., "Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk," Phys. Rev. E **70**, 045101 (2004).
[22] Sethuraman, J. and Sethuraman, S., "Connections between Bernoulli strings and random permutations," in *The Legacy of Alladi Ramakrishnan in the Mathematical Sciences* (Springer-Verlag, New York, 2010).
[23] Wu, L., Qi, Y., and Yang, J., "Asymptotics for dependent Bernoulli random variables," Stat. Probab. Lett. **82**(3), 455–463 (2012).
[24] Zelterman, D., *Discrete Distributions: Applications in the Health Sciences* (John Wiley & Sons, New York, 2004).
[25] Zhang, Y. and Zhang, L., "On the almost sure invariance principle for dependent Bernoulli random variables," Stat. Probab. Lett. **107**, 264–271 (2015).