Renewal Model for Dependent Binary Sequences

Marco Zamparo[1]

Abstract
We suggest to construct infinite stochastic binary sequences by associating one of the two symbols of the sequence with the renewal times of an underlying renewal process. Focusing on stationary binary sequences corresponding to delayed renewal processes, we investigate correlations and the ability of the model to implement a prescribed autocovariance structure, showing that a large variety of subexponential decay of correlations can be accounted for. In particular, robustness and efficiency of the method are tested by generating binary sequences with polynomial and stretched-exponential decay of correlations. Moreover, to justify the maximum entropy principle for model selection, an asymptotic equipartition property for typical sequences that naturally leads to the Shannon entropy of the waiting time distribution is demonstrated. To support the comparison of the theory with data, a law of large numbers and a central limit theorem are established for the time average of general observables.

1 Introduction

Binary processes arise in natural and social sciences any time a phenomenon is dichotomized according to the occurrence or non-occurrence of a property of interest. Modeling, statistical inference from data, and generation of dependent binary sequences, both finite and infinite, have thus received a lot of attention in a number of disciplines, such as statistical physics [1], system biology [2], computational neuroscience [3], actuarial and financial sciences [4], and machine learning [5], among others. While finite sequences are generally associated with graphical models [6], infinite sequences are usually understood as time series.

✉ Marco Zamparo
marco.zamparo@uniba.it

1 Dipartimento di Fisica, Università degli Studi di Bari and INFN, Sezione di Bari, via Amendola 173, 70126 Bari, Italy

Modeling and generating binary sequences with prescribed correlations is one of the major problems to be faced [1]. For finite sequences, the solution to the problem is in principle represented by the Ising model, which is an exponential family able to reproduce all inputs in (the interior of) the correlation polytope [6]. For infinite sequences, the problem of constructing binary variables with given correlations is more challenging since there is no universal framework to refer to. Indeed, finite-dimensional distributions structured according to the Ising model are generally not consistent under marginalization, so that they cannot be regarded as children of a common probability measure associated with a stochastic process. On the other hand, the Ising model is the only known parametric family that can cover all correlations of a finite number of binary variables. For this reason, in order to describe and generate infinite binary sequences, researchers have devised several approaches, each with its own advantages and disadvantages, which can be grouped in autoregressive models [7–14], latent factor models [15–17], and mixed models which combine autoregression with latent factors [7,20]. Autoregressive models are Markov chains with the property that the current probability of a symbol conditional on the past history is determined by a linear function of the previous outcomes [7], in a number corresponding to the order of the chain, linearity being postulated to not have explosion of parameters. These type of models can produce any exponential decay of correlations, whereas subexponential decays are off limits due to Markovianity and finite state space [21]. Latent factor models rely on an underlying latent process, and the first proposal to generate an infinite dependent binary sequence was to clip a latent Gaussian process at a fixed level [15]. Different latent factor approaches are represented by mixture models whereby the binary symbols are drawn independently with law determined by the realization of an underlying process, such as a latent Gaussian process [16] or a latent binary process [17]. Latent factor approaches can introduce a long-range dependence between the binary variables and can thus describe both exponential and subexponential decays of correlations, such as polynomial decay. However, the analysis of clipping performed in [18] has revealed that there are serious restrictions to the type of correlations that can be produced. In fact, decay rates, in the case of exponential decay, and decay exponents, in the case of polynomial decay, cannot be smaller than a certain minimum threshold. Similar restrictions have been found in mixture models with latent binary inputs [17]. These restrictions have been weakened in [19] by means of an algorithm that by iterating the latent factor model of [17], which transforms the latent binary input in a binary output, adds dependence to an initial uncorrelated binary sequence step by step. Other algorithms to generate dependent binary sequences and their demonstration through numerical experiments are discussed in [22–25].

In this paper we suggest and investigate a new latent factor approach for infinite binary sequences that relies on renewal processes [26]. Our proposal is to associate one of the two symbols of the sequence with the renewal times of a discrete-time delayed renewal process, and the other symbol with all other times, thus defining a regenerative phenomenon a la Kingman [27]. Delay makes it possible to construct stationary binary sequences. This way, we can offer a flexible and powerful model that is under full mathematical control and that is able to implement a large variety of subexponential prescriptions for the correlations, from polynomial to stretched-exponential decays as we shall demonstrate. Furthermore, reflecting an underlying renewal structure, generation of the process in numerical experiments is immediate.

The paper is organized as follows. In Sect. 2 we introduce the model and discuss its Markovian limit. Section 3 is devoted to correlations. In this section we prove the fundamental property of the model that binary symbols that follows a renewal time are independent of previous symbols and we show that the autocovariance solves a renewal equation.

Then we investigate the direct problem of determining the asymptotics of correlations from the waiting time distribution and the inverse problem of associating a renewal structure to a prescribed autocovariance for binary symbols. In Sect. 4 we prove limit theorems for the probability of typical sequences and for the time averages of general observables. In particular, we demonstrate an asymptotic equipartition property that naturally introduces the Shannon entropy of the waiting time distribution and leads to the maximum entropy principle for model selection [28,29]. Then we provide a law of large numbers and a central limit theorem for the normal fluctuations of empirical means, and we apply them to estimation of the waiting time distribution and of the autocovariance from data. These results cannot be deduced from the standard limit theory of regenerative processes with independent cycles [26] and require new arguments. Conclusions are drawn in Sect. 5. The most technical mathematical proofs are reported in the appendices in order to not interrupt the flow of the presentation.

2 Definition of the Model

Let on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ be given positive integer-valued random variables $S_1, S_2, \ldots$. We construct a binary stochastic sequence $X := \{X_t\}_{t \geq 1}$ with entries valued in $\mathbb{N}_2 := \{0, 1\}$ by supposing that the variable S_n is the *waiting time* for the nth symbol 1 of the sequence, which occurs at the *renewal time* $T_n := S_1 + \cdots + S_n$. Thus, $X_t := 1$ if $t \in \{T_n\}_{n \geq 1}$ and $X_t := 0$ otherwise for each $t \geq 1$. Our fundamental assumptions are (*i*) that waiting times are independent and (*ii*) that $S_2, S_3, \ldots$ have the same distribution, which could differ from the distribution of S_1. Under these assumptions, the sequence $\{T_n\}_{n \geq 1}$ is a *renewal process*, *delayed* if S_1 is not distributed as S_2. We refer to [26] for basics of renewal theory. The reason for letting S_1 behave differently from the other waiting times is that its distribution can be chosen in such a way that the sequence X is stationary, as explained by the following theorem which is proved in Appendix A. We recall that the process X is *stationary* if $\{X_{t+1}\}_{t \geq 1}$ is distributed as $\{X_t\}_{t \geq 1}$, where $\{X_t\}_{t \geq 1}$ is understood as a random element in the set $\mathbb{N}_2^\infty$ of all binary sequences equipped with the cylinder σ-field $\mathscr{B}$. Hereafter, $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$.

Theorem 2.1 *The binary sequence X is stationary if and only if $\mathbb{E}[S_2] < \infty$ and for every $s \geq 1$*

$$\mathbb{E}[S_2] \, \mathbb{P}[S_1 = s] = \mathbb{P}[S_2 \geq s].$$

Theorem 2.1 recovers well-known conditions for the delayed renewal chain $\{T_n\}_{n \geq 1}$ to be stationary, that means that the statistical properties of the number of renewals in a given temporal window are invariant with respect to time shifts [26]. Under the hypothesis of stationarity, which is now made, the only input of the model X is the distribution of the waiting time S_2. Depending on the need, we refer to its density $p := \mathbb{P}[S_2 = \cdot]$ or to its tail $Q := \mathbb{P}[S_2 > \cdot]$. Thus, according to Theorem 2.1, we suppose that $\mu := \mathbb{E}[S_2] = \sum_{s \geq 1} s p(s) = \sum_{t \geq 0} Q(t) < \infty$ and that $\mathbb{P}[S_1 = s] = (1/\mu) \sum_{n \geq s} p(n)$ for every $s \geq 1$. We notice that $\mathbb{E}[X_1] = \mathbb{P}[T_1 = 1] = \mathbb{P}[S_1 = 1] = 1/\mu$.

The analysis of the finite-dimensional distributions reveals that if the sequence X is stationary, then it is *time-reversible* in the sense that $(X_t, \ldots, X_2, X_1)$ is distributed as $(X_1, X_2, \ldots, X_t)$ for all t. The finite-dimensional distributions of X are provided by the next proposition, which is demonstrated in Appendix B. We make the usual conventions that an empty sum is assumed to be 0 and an empty product is assumed to be 1.

Proposition 2.1 *For any integer $t \geq 1$ and numbers $x_1, x_2, \ldots, x_t$ in $\mathbb{N}_2$*

$$\pi_t(x_1, x_2, \ldots, x_t) := \mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t]$$

$$= \frac{1}{\mu} \left[\sum_{s \geq t} Q(s) \right]^{\prod_{n=1}^{t}(1-x_n)} \prod_{i=1}^{t} [Q(i-1)]^{\prod_{n=1}^{i-1}(1-x_n)x_i}$$

$$\cdot \prod_{i=1}^{t-1} \prod_{j=i+1}^{t} [p(j-i)]^{x_i \prod_{n=i+1}^{j-1}(1-x_n)x_j} \prod_{j=1}^{t} [Q(t-j)]^{x_j \prod_{n=j+1}^{t}(1-x_n)}$$

$$= \pi_t(x_t, \ldots, x_2, x_1)$$

with $0^0 := 1$.

Proposition 2.1 can be used to compute the conditional probability of a symbol given the past and allows us to get some insight into the dependence structure of the stochastic sequence X. Let us denote by $l_t(x_1, \ldots, x_t)$ the position of the first symbol 1 in the binary string $(x_1, \ldots, x_t)$: $l_t(x_1, \ldots, x_t) := l$ if $x_l = 1$ and $x_1 = \cdots = x_{l-1} = 0$ for some $l \leq t$, and $l_t(x_1, \ldots, x_t) := \infty$ if $x_1 = \cdots = x_t = 0$. Proposition 2.1 shows through easy manipulations that for any positive integer t and binary numbers $x_1, \ldots, x_t$

$$\mathbb{P}[X_{t+1} = 1 | X_t = x_1, \ldots, X_1 = x_t] = \begin{cases} \dfrac{Q(t)}{\sum_{s \geq t} Q(s)} & \text{if } l_t(x_1, \ldots, x_t) = \infty; \\ \dfrac{p(l)}{Q(l-1)} & \text{if } l_t(x_1, \ldots, x_t) =: l < \infty \end{cases} \tag{2.1}$$

provided that $\mathbb{P}[X_1 = x_t, \ldots, X_t = x_1] > 0$. Pay attention to the inverse order of $x_1, \ldots, x_t$ in the conditioning event. Formula (2.1) draws a link between our model and the so-called "stochastic chains with memory of variable length" [30]. In this class of models, which can be built on a larger alphabet than $\mathbb{N}_2$, the number of variables involved in the conditioning event is determined by a "context length function", which itself depends on the past variables. The context length function of our model is exactly the function l_t that maps any binary string $(x_1, \ldots, x_t)$ in $l_t(x_1, \ldots, x_t)$. Since the function l_t is unbounded as the time t goes on, our model constitutes a stochastic chain having in general memory of unbounded variable length.

In spite of an unbounded context length, the process X is a Markov chain for certain particular waiting time distributions p. The sequence X is a *Markov chain* of order $M \geq 1$ if the conditional probability of a symbol given the past depends only on the state of the last M variables. We can include sequences of i.i.d. random variables in this definition by letting M to take the value zero. The requirement that the conditional probability (2.1) is independent of $x_{M+1}, \ldots, x_t$ for Markovianity of order $M \geq 0$ results in the following corollary of Proposition 2.1, whose proof is reported in Appendix C.

Corollary 2.1 *The binary sequence X is a Markov chain of order $M \geq 0$ if and only if there exists a real number $\lambda \in [0, 1)$ such that $p(M + s + 1) = \lambda^s p(M + 1)$ for all $s \geq 1$.*

3 Correlations

The binary sequence $X := \{X_t\}_{t \geq 1}$ is a *regenerative process* with regeneration points $\{T_n\}_{n \geq 1}$ and independent cycles [26]. This means that the past and the future of each renewal event, i.e. of each symbol 1, are independent. Such conditional independence is stated by the following proposition, whose proof is provided in Appendix D.

Proposition 3.1 *For every positive integers $s \leq t$ and binary numbers $x_1, \ldots, x_t$*

$$\mathbb{P}[X_1 = x_1, \ldots, X_t = x_t | X_s = 1] = \mathbb{P}[X_1 = x_1, \ldots, X_s = x_s | X_s = 1]$$
$$\cdot \, \mathbb{P}[X_s = x_s, \ldots, X_t = x_t | X_s = 1].$$

Conditional independence associated with renewals is exploited in this section to study correlations. The *covariance* of two random variables Y and Z on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is denoted by $\mathrm{cov}[Y, Z]$:

$$\mathrm{cov}[Y, Z] := \mathbb{E}\big[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])\big] = \mathbb{E}[YZ] - \mathbb{E}[Y]\,\mathbb{E}[Z].$$

For every $t \geq 0$, the *autocovariance* $\rho_t := \mathrm{cov}[X_1, X_{t+1}]$ reads $\rho_t = c_t - c_0^2$ with

$$c_t := \mathbb{E}[X_1 X_{t+1}].$$

Notice that $c_0 = \mathbb{E}[X_1] = 1/\mu$. The next proposition, which is based on Proposition 3.1 and is proved in Appendix E, shows that the sequence $c := \{c_t\}_{t \geq 0}$ solves a renewal equation with waiting time distribution p. As a consequence, the renewal theorem [31] gives $\lim_{t \uparrow \infty} c_t = c_0/\mu = c_0^2$, namely $\lim_{t \uparrow \infty} \rho_t = 0$, provided that p is aperiodic. The probability distribution p is *aperiodic* if there is no proper sublattice of $\{1, 2, \ldots\}$ containing its support.

Proposition 3.2 *For every integer $t \geq 1$*

$$c_t = \sum_{s=1}^{t} p(s)\, c_{t-s}.$$

By resorting to the literature on the renewal equation, we investigate the direct problem of determining the asymptotics of the sequence c, and hence of the autocovariance, from a given waiting time distribution p, as well as the inverse problem of finding conditions on a prescribed c such that c solves a renewal equation with some probability distribution p of finite mean. The latter aims to answer the question of which autocovariance structures can be reproduced by our model. Before that, we want to point out that the autocovariance of the binary sequence X describes the time dependence of any temporal correlations, as shown by the following technical lemma that explores temporal correlations of general observables. The lemma is proved in Appendix F and will be used in Sect. 4 to address the mixing properties of X.

Lemma 3.1 *Fix positive integers m and n and let f and g be two real functions on $\mathbb{N}_2^m$ and $\mathbb{N}_2^n$, respectively, such that $f(0, \ldots, 0) = g(0, \ldots, 0) = 0$. Set $Z_t := g(X_t, \ldots, X_{t+n-1})$ for $t \geq 1$. Then, for all $t \geq 1$*

$$\mathrm{cov}\big[f(X_1, \ldots, X_m), Z_{m+t}\big] = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{i,j}(t)\, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=i\}} f(X_m, \ldots, X_1)]}{\mathbb{P}[S_1 = i]}\, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=j\}} Z_1]}{\mathbb{P}[S_1 = j]},$$

where for $i \geq 1$ and $j \geq 1$

$$C_{i,j}(t) := \sum_{u=1}^{i} \sum_{v=1}^{j} p(i - u)\, \rho_{u+t+v-2}\, p(j - v)$$

with $p(0) := -1$.

3.1 Autocorrelation: Direct Problem

Let us study the asymptotics of the autocovariance for a given waiting time distribution. Pure exponential decay of correlations can be described by Markov chains as identified by Corollary 2.1 and is somehow trivial. We shall touch on the exponential behavior of the auto-covariance when dealing with the inverse problem. Here we focus on subexponential decays that account for long-range dependence that cannot be explained by Markov processes. A natural setting for subexponentiality in renewal theory was given in [32]. Let the symbol $\sim$ denotes asymptotic equivalence for sequences: $a_t \sim b_t$ means $\lim_{t \uparrow \infty} a_t/b_t = 1$. Following [32], we say that a positive real sequence $a := \{a_t\}_{t \geq 0}$ belongs to the class $\mathscr{S}$ of *subexponential sequences* if $A := \sum_{t \geq 0} a_t < \infty$, $a_{t+1} \sim a_t$, and $\sum_{n=0}^{t} a_n a_{t-n} \sim 2Aa_t$. The requirement $a_{t+1} \sim a_t$ prevents a from growing exponentially, thus justifying the terminology "subexponential". The asymptotic behavior of the autocovariance ρ_t can be characterized in general when there exists $\lambda \in (0, 1]$ such that $\{\lambda^{-t} Q(t)\}_{t \geq 0} \in \mathscr{S}$, namely when the tail probability Q has exponential rate λ and a subexponential correction that forms a sequence in $\mathscr{S}$. The case $\lambda = 1$ corresponds to a pure subexponential behavior. In fact, Theorem 3.2 of [33] for the rate of convergence of renewal sequences gives the following result.

Theorem 3.1 *Assume that p is aperiodic and that $\{\lambda^{-t} Q(t)\}_{t \geq 0} \in \mathscr{S}$ for some $\lambda \in (0, 1]$. Then*

$$\rho_t \sim \frac{\sum_{n > t} Q(n)}{\mu \left[\sum_{n \geq 0} \lambda^{-n} Q(n) \right]^2}.$$

Subexponential behaviors that find wide application are polynomial decays, which fall under the umbrella of regular variation, and Weibull-type decays represented by stretched exponentials. We now suppose $\lambda = 1$ and discuss these decays in some detail. We stress that the necessary condition $\sum_{t \geq 0} Q(t) < \infty$ for the sequence $\{Q(t)\}_{t \geq 0}$ to belong to $\mathscr{S}$ is satisfied since $\sum_{t \geq 0} Q(t) = \mu < \infty$ by the hypothesis of stationarity.

3.1.1 Polynomial Decay

A positive sequence $a := \{a_t\}_{t \geq 0}$ is *regularly varying* if there exists an index $\alpha \in \mathbb{R}$ and a slowly varying function ℓ such that $a_t \sim t^\alpha \ell(t)$. A real measurable function ℓ is *slowly varying* if it is positive on a neighborhood of infinity, say (τ, ∞) with some $\tau > 0$, and satisfies the scale-invariance property $\lim_{z \uparrow \infty} \ell(\eta z)/\ell(z) = 1$ for any number $\eta > 0$. Trivially, a measurable function with a finite positive limit at infinity is slowly varying. The simplest non-trivial example is represented by the logarithm. We refer to [34] for the theory of slow and regular variation. We stress that a slowly varying function ℓ is dominated by polynomials in the sense that $\lim_{z \uparrow \infty} z^\gamma \ell(z) = \infty$ and $\lim_{z \uparrow \infty} z^{-\gamma} \ell(z) = 0$ for all $\gamma > 0$ according to Proposition 1.3.6 of [34]. The uniform convergence theorem [34] states that the scale-invariance property of slowly varying functions actually is uniform for η that belongs to each compact set in $(0, \infty)$. This fact can be used to show that $a_{t+1} \sim a_t$ if a is a regularly varying sequence. Combined with the dominated convergence theorem, it also shows that $\sum_{n=0}^{t} a_n a_{t-n} \sim 2 \sum_{n=0}^{\lfloor t/2 \rfloor} a_n a_{t-n} \sim 2Aa_t$ when $A := \sum_{t \geq 0} a_t < \infty$, $\lfloor t/2 \rfloor$ denoting the integer part of $t/2$. Thus, any summable regularly varying sequence is an element of $\mathscr{S}$. Summability imposes the restriction $\alpha \leq -1$ to the index.

These arguments show that if $Q(t) \sim t^{-\gamma-1} \ell(t)$ with an exponent $\gamma > 0$ and an arbitrary slowly varying function ℓ, then $\{Q(t)\}_{t \geq 0} \in \mathscr{S}$. In such case we have the asymptotic

equivalence $\sum_{n>t} Q(n) \sim (1/\gamma)t^{-\gamma}\ell(t)$ by Proposition 1.5.10 of [34]. Thus, we get the following corollary of Theorem 3.1.

Corollary 3.1 *Assume that p is aperiodic and $Q(t) \sim t^{-\gamma-1}\ell(t)$ with an exponent $\gamma > 0$ and a slowly varying function ℓ. Then*

$$\rho_t \sim \frac{\ell(t)}{\gamma\mu^3 t^\gamma}.$$

In contrast to the latent factor models analyzed in [17] and [18], which can account for polynomial decay of correlations but not for too small exponents, Corollary 3.1 of Theorem 3.1 shows that a renewal structure is able to describe polynomial decays of the autocovariance with any exponent $\gamma > 0$. Actually, the hypothesis $\gamma > 0$ is not necessary and we can also have $\gamma = 0$, so that the autocovariance decays slower than any polynomial, but in this case the asymptotic behavior of $\{\sum_{n>t} Q(n)\}_{t\geq 0}$ cannot be resolved in general. Notice that summability of $\{Q(t)\}_{t\geq 0}$ when $\gamma = 0$ imposes restrictions on ℓ. For instance, if $Q(t) \sim t^{-1}(\ln t)^{-\beta-1}$ with a number $\beta > 0$, then $\{Q(t)\}_{t\geq 0} \in \mathscr{S}$ and Theorem 3.1 gives $\rho_t \sim (1/\mu^3)\sum_{n>t} Q(n) \sim (1/\beta\mu^3)(\ln t)^{-\beta}$.

3.1.2 Stretched-Exponential Decay

In [32] the following sufficient condition for subexponentiality was proposed. Let h be a continuously differentiable real function on a neighborhood of infinity, say (τ, ∞) with some $\tau > 0$, such that its derivative h' enjoys the properties that $-z^2 h'(z)$ is increasing to infinity with respect to z and that $\int_\tau^\infty e^{\frac{1}{2}z^2 h'(z)}dz < \infty$. Then, a positive sequence $a := \{a_t\}_{t\geq 0}$ such that $A := \sum_{t\geq 0} a_t < \infty$ and $a_{t+1} \sim a_t \sim e^{-th(t)}$ satisfies $\sum_{n=0}^t a_n a_{t-n} \sim 2Aa_t$, and hence belongs to $\mathscr{S}$. We have used this criterion to investigate the asymptotic behavior of the autocovariance when $Q(t) \sim e^{-t^\beta \ell(t)}$ with a stretching exponent $\beta \in (0, 1)$ and some function ℓ. The following corollary of Theorem 3.1, which is proved in Appendix G, gives sufficient conditions on ℓ that imply $\{Q(t)\}_{t\geq 0} \in \mathscr{S}$. We point out that ℓ is slowly varying under those conditions.

Corollary 3.2 *Assume that p is aperiodic and $Q(t) \sim e^{-t^\beta \ell(t)}$ with a stretching exponent $\beta \in (0, 1)$ and a twice continuously differentiable positive function ℓ on a neighborhood of infinity that satisfies $\lim_{z\uparrow\infty} z\ell'(z)/\ell(z) = 0$ and $\lim_{z\uparrow\infty} z^2\ell''(z)/\ell(z)$ exists. Then*

$$\rho_t \sim \frac{t^{1-\beta}}{\mu^3 \beta\ell(t)}e^{-t^\beta \ell(t)}.$$

To conclude, we observe that the hypothesis $\beta < 1$ can be relaxed in favor of $\beta = 1$ to come fairly close to exponential decay of correlations while staying in the framework of subexponential sequences. If for example $Q(t) \sim e^{-t(\ln t)^{-\gamma}}$ with some number $\gamma > 0$, then $Q(t + 1) \sim Q(t) \sim e^{-th(t)}$ with $h(z) := (\ln z)^{-\gamma}$. This function h satisfies the above sufficient condition for subexponentiality, so that $\rho_t \sim (1/\mu^3)\sum_{n>t} Q(n) \sim (1/\mu^3)(\ln t)^\gamma e^{-t(\ln t)^{-\gamma}}$ by Theorem 3.1.

3.2 Autocorrelation: Inverse Problem

Let us now investigate the possibility to implement a prescribed autocovariance. Here the focus is on short time scales since Theorem 3.1 and its Corollaries 3.1 and 3.2 already settle

the issue on long time scales, demonstrating that a large class of asymptotic prescriptions can be obtained. We want to understand conditions on a non-negative sequence $c := \{c_t\}_{t\geq 0}$ under which there exists a waiting time distribution p of finite mean whose associated stationary binary sequence $X := \{X_t\}_{t\geq 1}$ satisfies $c_t = \mathbb{E}[X_1 X_{t+1}]$ for all $t \geq 0$. For simplicity, we assume that $c_t > 0$ for every $t \geq 0$. In the light of Theorem 2.1 and Proposition 3.2, this is tantamount to ask under which conditions on c there exists a probability distribution p with the properties $\sum_{s\geq 1} s p(s) = 1/c_0$ and $\sum_{s=1}^{t} p(s) c_{t-s} = c_t$ for all $t \geq 1$. Such a p, if any, is uniquely defined by the renewal equation and meets the requirement $\sum_{s\geq 1} s p(s) = 1/c_0$ if and only if $\lim_{t\uparrow\infty} c_t = c_0^2$. In fact, if p is a probability distribution such that $\sum_{s=1}^{t} p(s) c_{t-s} = c_t$ for all $t \geq 1$, then it is aperiodic since $p(1) = c_1/c_0 > 0$. This way, the renewal theorem [31] gives $\lim_{t\uparrow\infty} c_t = c_0/\sum_{s\geq 1} s p(s)$ if $\sum_{s\geq 1} s p(s) < \infty$ and $\lim_{t\uparrow\infty} c_t = 0$ if $\sum_{s\geq 1} s p(s) = \infty$.

Finding the minimal conditions on the sequence c for the existence of an associated waiting time distribution p is a difficult task. We stress that the problem consists in determining whether or not the function p that solves the problem $\sum_{s=1}^{t} p(s) c_{t-s} = c_t$ for every $t \geq 1$ is non-negative and sums to 1. However, there is a sufficient condition that covers many applications. A sequence $c := \{c_t\}_{t\geq 0}$ is a *Kaluza sequence* [27] if $c_t > 0$ and $c_{t-1} c_{t+1} \geq c_t^2$ for all $t \geq 1$. It follows that $c_0 > 0$. The following theorem states that the hypothesis that c is a Kaluza sequence such that $\lim_{t\uparrow\infty} c_t = c_0^2$ guarantees the existence of an associated waiting time distribution of finite mean. The proof is provided in Appendix H.

Theorem 3.2 *Let $c := \{c_t\}_{t\geq 0}$ be a Kaluza sequence such that $\lim_{t\uparrow\infty} c_t = c_0^2$. Then, there exists a unique waiting time distribution p with the properties $\sum_{s\geq 1} s p(s) = 1/c_0$ and $\sum_{s=1}^{t} p(s) c_{t-s} = c_t$ for all $t \geq 1$. As a consequence, the stationary binary sequence $X := \{X_t\}_{t\geq 1}$ associated with p satisfies $c_t = \mathbb{E}[X_1 X_{t+1}]$ for all $t \geq 0$.*

We point out that Theorem 3.2 of [33] offers an inverse of Theorem 3.1: if $c := \{c_t\}_{t\geq 0}$ is a Kaluza sequence such that $\lim_{t\uparrow\infty} c_t = c_0^2$ and $\{c_{t+1} - c_t\}_{t\geq 0} \in \mathscr{S}$, then there exists a unique associated waiting time distribution p whose tail Q enjoys the property $(1/\mu^3) \sum_{n>t} Q(n) \sim \rho_t := c_t - c_0^2$.

A practical criterion to recognize a Kaluza sequence that puts the emphasis on the auto-covariance is the following. Consider a sequence $c := \{c_t\}_{t\geq 0}$ defined by $c_0 := \xi$ and $c_t := \xi^2 + m e^{-\phi(t)}$ for $t \geq 1$ with constants $\xi \in (0, 1]$ and $m \in [0, \xi(1-\xi)]$ and a concave function ϕ such that $\phi(0) = 0$ and $\lim_{z\uparrow\infty} \phi(z) = \infty$. Then, c is a Kaluza sequence and $\lim_{t\uparrow\infty} c_t = c_0^2$. Indeed, $c_0 = \xi^2 + \xi(1-\xi) \geq \xi^2 + m e^{-\phi(0)}$, so that for each $t \geq 1$

$$
\begin{aligned}
c_{t-1} c_{t+1} - c_t^2 &\geq [\xi^2 + m e^{-\phi(t-1)}][\xi^2 + m e^{-\phi(t+1)}] - [\xi^2 + m e^{-\phi(t)}]^2 \\
&= m\xi^2 [e^{-\phi(t-1)} + e^{-\phi(t+1)} - 2e^{-\phi(t)}] \\
&\quad + m^2 e^{-2\phi(t)} [e^{2\phi(t)-\phi(t-1)-\phi(t+1)} - 1].
\end{aligned}
$$

This way, the concavity of ϕ and the consequent convexity of $e^{-\phi}$ give $c_{t-1} c_{t+1} - c_t^2 \geq 0$ for all $t \geq 1$. We have thus proved the following corollary of Theorem 3.2.

Corollary 3.3 *Let $\xi \in (0, 1]$ and $m \in [0, \xi(1-\xi)]$ be two constants and let ϕ be a concave function such that $\phi(0) = 0$ and $\lim_{z\uparrow\infty} \phi(z) = \infty$. Then, there exists a unique waiting time distribution p of finite mean whose associated stationary binary sequence $X := \{X_t\}_{t\geq 1}$ satisfies $\mathbb{E}[X_1] = \xi$ and $\rho_t := \mathrm{cov}[X_1, X_{t+1}] = m e^{-\phi(t)}$ for all $t \geq 1$.*

We appeal to Corollary 3.3 to draw some examples. Our model can implement the autocovariances $\rho_t = m(1+t)^{-\gamma}$ and $\rho_t = m e^{-\kappa t^{\beta}}$ for $t \geq 1$ with any real numbers
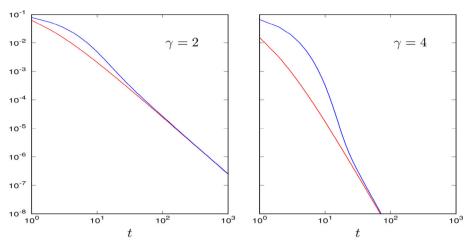
Fig. 1 Autocovariance ρ_t (red) and $(1/\mu^3)\sum_{n>t} Q(n)$ (blue) versus t for the inverse problem with $c_t = 1/4 + (1/4)(1+t)^{-\gamma}$ for $t \geq 0$ with $\gamma = 2$ (left) and $\gamma = 4$ (right) (Color figure online)
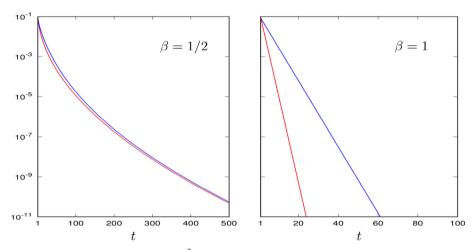

Fig. 2 Autocovariance ρ_t (red) and $(1/\mu^3)\sum_{n>t} Q(n)$ (blue) versus t for the inverse problem with $c_t = 1/4 + (1/4)e^{-t^\beta}$ for $t \geq 0$ with $\beta = 1/2$ (left) and $\beta = 1$ (right) (Color figure online)

$\xi \in (0, 1]$, $m \in [0, \xi(1-\xi)]$, $\gamma > 0$, $\kappa > 0$, and $\beta \in (0, 1]$. Figure 1 compares ρ_t with $(1/\mu^3)\sum_{n>t} Q(n)$ for the two inverse problems corresponding to the polynomial correlation structures $c_t = 1/4 + (1/4)(1+t)^{-\gamma}$ for $t \geq 0$ with exponents $\gamma = 2$ and $\gamma = 4$, respectively. We see that the earlier mentioned asymptotic equivalence $\rho_t \sim (1/\mu^3)\sum_{n>t} Q(n)$ is confirmed. Figure 2 shows the same comparison for the two stretched-exponential cases $c_t = 1/4 + (1/4)e^{-t^\beta}$ for $t \geq 0$ with exponents $\beta = 1/2$ and $\beta = 1$, respectively. This time, the asymptotic equivalence $\rho_t \sim (1/\mu^3)\sum_{n>t} Q(n)$ is confirmed only for the subexponential case $\beta = 1/2$. The exponential case $\beta = 1$ is solved explicitly by the formulas $\rho_t = (1/4)e^{-t}$ and $(1/\mu^3)\sum_{n>t} Q(n) = (1/8)(\frac{1+e^{-1}}{2})^t$ for all $t \geq 0$, which shows that the decay rates of the autocovariance and of the associated waiting time distribution are different.

4 Limit Theorems

There are a number of limit theorems for the sequence $X := \{X_t\}_{t \geq 1}$ that follow from an underlying ergodic property. In this section we discuss some of these limit theorems, proving at first ergodicity of a related dynamical system. We refer to [35] for basics of ergodic theory. Recalling that $\mathscr{B}$ denotes the σ-field on $\mathbb{N}_2^\infty$ generated by the cylinder subsets, it is now convenient to introduce the probability measure $\mathbb{P}_o[\,\cdot\,] := \mathbb{P}[X \in \cdot\,]$ on $\mathscr{B}$ and the probability space $(\mathbb{N}_2^\infty, \mathscr{B}, \mathbb{P}_o)$. We can deduce almost sure convergence in $(\Omega, \mathscr{F}, \mathbb{P})$ from almost sure convergence in $(\mathbb{N}_2^\infty, \mathscr{B}, \mathbb{P}_o)$. In fact, since almost sure convergence is defined only in terms of probability distributions [36], if $G, G_1, G_2, \ldots$ are real $\mathscr{B}$-measurable functions on $\mathbb{N}_2^\infty$ such that $\lim_{n \uparrow \infty} G_n(x) = G(x)$ for $\mathbb{P}_o$-a.e. x, then $\lim_{t \uparrow \infty} G_t(X) = G(X)$ $\mathbb{P}$-a.s. The same can be said for convergence in mean. The reason to deal with the new probability space $(\mathbb{N}_2^\infty, \mathscr{B}, \mathbb{P}_o)$ is that it can be endowed with a measure-preserving transformation. Such transformation is the left-shift operator $\mathcal{T}$ that maps any binary sequence $x := \{x_t\}_{t \geq 1} \in \mathbb{N}_2^\infty$ in $\mathcal{T}x := \{x_{t+1}\}_{t \geq 1}$. The operator $\mathcal{T}$ is measurable and, due to stationary of X, preserves measures, namely $\mathbb{P}_o[\mathcal{T}^{-1}\mathcal{B}] = \mathbb{P}_o[\mathcal{B}]$ for each $\mathcal{B} \in \mathscr{B}$. Indeed, if $\mathcal{B}$ is an element of $\mathscr{B}$, then $\mathbb{P}_o[\mathcal{T}^{-1}\mathcal{B}] = \mathbb{P}[\mathcal{T}X \in \mathcal{B}] = \mathbb{P}[\{X_{t+1}\}_{t \geq 1} \in \mathcal{B}] = \mathbb{P}[\{X_t\}_{t \geq 1} \in \mathcal{B}] = \mathbb{P}_o[\mathcal{B}]$. The transformation $\mathcal{T}$ is *strong-mixing* in the sense of ergodic theory, as stated by the following lemma that in proved in Appendix I. The proof relies on Lemma 3.1.

Lemma 4.1 *Assume that p is aperiodic. Then, for all $\mathcal{A}$ and $\mathcal{B}$ in $\mathscr{B}$*

$$\lim_{t \uparrow \infty} \mathbb{P}_o[\mathcal{A} \cap \mathcal{T}^{-t}\mathcal{B}] = \mathbb{P}_o[\mathcal{A}] \, \mathbb{P}_o[\mathcal{B}].$$

Due to strong-mixing, the transformation $\mathcal{T}$ is *ergodic* according to Corollary 1.14.2 of [35], which means that the only members $\mathcal{B}$ of $\mathscr{B}$ with $\mathcal{T}^{-1}\mathcal{B} = \mathcal{B}$ satisfy $\mathbb{P}_o[\mathcal{B}] = 0$ or $\mathbb{P}_o[\mathcal{B}] = 1$. We use ergodicity to demonstrate an asymptotic equipartition property, which justifies the principle of maximum entropy for model selection, and to investigate the behavior of empirical means.

4.1 Asymptotic Equipartition Property

Description of data requires to select a statistical model, that is a waiting time distribution p once our framework is considered. One tool for model selection is the *maximum entropy principle* [28,29], which would amount to pick the probability distribution p of finite mean that meets certain moment constraints representing the available information and that maximizes the Shannon entropy. The *Shannon entropy* $\mathcal{H}(p)$ of p is defined by

$$\mathcal{H}(p) := -\sum_{s \geq 1} p(s) \ln p(s)$$

with $0 \ln 0 := 0$. We point out that $\mathcal{H}(p) \leq \mu \ln \mu + (1 - \mu) \ln(\mu - 1) < \infty$ whenever $\mu := \sum_{s \geq 1} s p(s) < \infty$ since $-p(s) \ln p(s) \leq \mu^{-s}(\mu - 1)^{s-1} - p(s) - p(s) \ln[\mu^{-s}(\mu - 1)^{s-1}]$ for all $s \geq 1$ by concavity. Actually, the largest value of the entropy is $\mu \ln \mu + (1 - \mu) \ln(\mu - 1)$, which is attained if and only if $p(s) = \mu^{-s}(\mu - 1)^{s-1}$ for every s. The Shannon entropy can be derived axiomatically as a measure of uncertainty in the outcomes of a random variable. Instead, in this section we show that the entropy $\mathcal{H}(p)$ of p naturally arises as the answer to the question "how many typical sequences are there?". In fact, we demonstrate that there are about $e^{(t/\mu)\mathcal{H}(p)}$ typical sequences of length t, each with probability about $e^{-(t/\mu)\mathcal{H}(p)}$. It follows that selecting the waiting time distribution that maximizes the entropy

means not excluding possible sequences arbitrarily. If the only available information is the mean μ, then the maximum entropy prescription is the waiting time distribution p defined by $p(s) = \mu^{-s}(\mu - 1)^{s-1}$ for all s. According to Corollary 2.1, the binary sequence X associated with such p is a sequence of i.i.d. binary random variables with mean $1/\mu$.

Let us formalize and explain the above statements. Together with the Shannon entropy of the waiting time distribution, we consider for each $t \geq 1$ the Shannon entropy $\mathcal{H}(\pi_t)$ of the finite-dimensional distribution π_t of X:

$$\mathcal{H}(\pi_t) := - \sum_{x_1 \in \mathbb{N}_2} \cdots \sum_{x_t \in \mathbb{N}_2} \pi_t(x_1, \ldots, x_t) \ln \pi_t(x_1, \ldots, x_t).$$

Proposition 2.1 and stationary and reversibility of X give

$$\mathcal{H}(\pi_t) = \ln \mu - \mathbb{E}\left[\prod_{n=1}^{t}(1 - X_n) \right] \ln \sum_{s \geq t} Q(s) - 2 \sum_{s=1}^{t} \mathbb{E}\left[\prod_{n=1}^{s-1}(1 - X_n)X_s \right] \ln Q(s - 1)$$

$$- \sum_{s=1}^{t}(t - s)\, \mathbb{E}\left[X_1 \prod_{n=2}^{s}(1 - X_n)X_{s+1} \right] \ln p(s)$$

$$= \ln \mu - \frac{1}{\mu} \sum_{s \geq t} Q(s) \ln \sum_{s \geq t} Q(s) - \frac{2}{\mu} \sum_{s=1}^{t} Q(s - 1) \ln Q(s - 1)$$

$$- \frac{1}{\mu} \sum_{s=1}^{t}(t - s)\, p(s) \ln p(s),$$

where we have used the facts that $\mathbb{E}[\prod_{n=1}^{t}(1 - X_n)] = \mathbb{P}[S_1 > t]$, $\mathbb{E}[\prod_{n=1}^{s-1}(1 - X_n)X_s] = \mathbb{P}[S_1 = s]$, and $\mathbb{E}[X_1 \prod_{n=2}^{s}(1 - X_n)X_{s+1}] = \mathbb{P}[S_1 = 1, S_2 = s]$ for all s. After invoking the properties of Cesáro means and the dominated converge theorem, this formula shows that the *entropy rate* of the sequence X is

$$\lim_{t \uparrow \infty} \frac{\mathcal{H}(\pi_t)}{t} = \frac{\mathcal{H}(p)}{\mu}. \tag{4.1}$$

Then, under ergodicity of the left-shift operator $\mathcal{T}$, the Shannon–McMillan–Breiman theorem [37] yields for $\mathbb{P}_o$-a.e. x

$$\lim_{t \uparrow \infty} -\frac{1}{t} \ln \pi_t(x_1, \ldots, x_t) = \lim_{t \uparrow \infty} \frac{\mathcal{H}(\pi_t)}{t} = \frac{\mathcal{H}(p)}{\mu}.$$

These considerations, in combination with Lemma 4.1 and the possibility to deduce almost sure convergence in $(\Omega, \mathscr{F}, \mathbb{P})$ from almost sure convergence in $(\mathbb{N}_2^\infty, \mathscr{B}, \mathbb{P}_o)$, prove the following asymptotic equipartition property, which states that all long typical sequences of length t have roughly the same probability $e^{-(t/\mu)\mathcal{H}(p)}$.

Theorem 4.1 *Assume that p is aperiodic. Then*

$$\lim_{t \uparrow \infty} -\frac{\mu}{t} \ln \pi_t(X_1, \ldots, X_t) = \mathcal{H}(p) \quad \mathbb{P}\text{-a.s..}$$

Theorem 4.1 implies that there are about $e^{(t/\mu)\mathcal{H}(p)}$ typical sequences of length t. This fact is illustrated by the following corollary, whose proof is reported in Appendix J. A formal notion of typical set is needed here. Given a real number $\delta \in (0, 1)$, according to [37] we say that $\mathcal{X} \subseteq \mathbb{N}_2^t$ is a *typical set* of length $t \geq 1$ if $\mathbb{P}[(X_1, \ldots, X_t) \in \mathcal{X}] \geq 1 - \delta$. We denote by $|\mathcal{X}|$ the cardinality of a set $\mathcal{X}$.

Corollary 4.1 *Fix $\epsilon > 0$ and assume that p is aperiodic. The following conclusions hold for all sufficiently large t:*

  (i) *there exists a typical set $\mathcal{X}_o$ of length t such that $|\mathcal{X}_o| \leq e^{(t/\mu)[\mathcal{H}(p)+\epsilon]}$;*
 (ii) *any typical set $\mathcal{X}$ of length t satisfies $|\mathcal{X}| \geq e^{(t/\mu)[\mathcal{H}(p)-\epsilon]}$.*

4.2 Empirical Means

Applications to real data require to explain whether or not ensemble averages can be estimated by means of time averages, also known as empirical means. If G is a $\mathscr{B}$-measurable function on $\mathbb{N}_2^\infty$, then its *empirical mean* up to time $t \geq 1$ is the random variable

$$\frac{1}{t}\sum_{n=1}^{t} G(X_n, X_{n+1}, \ldots) = \frac{1}{t}\sum_{n=0}^{t-1} G(\mathcal{T}^n X).$$

Birkhoff ergodic theorem [35] tells us that if the left-shift operator $\mathcal{T}$ is ergodic and the expectation $\int_{\mathbb{N}_2^\infty} |G(x)|\,\mathbb{P}_o[dx] = \mathbb{E}[|G(X)|]$ is finite, then for $\mathbb{P}_o$-a.e. x

$$\lim_{t\uparrow\infty}\frac{1}{t}\sum_{n=0}^{t-1} G(\mathcal{T}^n x) = \int_{\mathbb{N}_2^\infty} G(x)\,\mathbb{P}_o[dx] = \mathbb{E}[G(X)].$$

The convergence also holds in mean by Corollary 1.14.1 of [35]. This way, Lemma 4.1 and the possibility to export convergences from $(\mathbb{N}_2^\infty, \mathscr{B}, \mathbb{P}_o)$ to $(\Omega, \mathscr{F}, \mathbb{P})$ result in the following law of large numbers, which gives a positive answer to the possibility of estimating ensemble averages with time averages. We stress that in most applications the observable G depends only on a finite number of variables, so that G is automatically $\mathscr{B}$-measurable and bounded.

Theorem 4.2 *Let G be a $\mathscr{B}$-measurable function on $\mathbb{N}_2^\infty$ such that $\mathbb{E}[|G(X)|] < \infty$. If p is aperiodic, then*

$$\lim_{t\uparrow\infty}\frac{1}{t}\sum_{n=0}^{t-1} G(\mathcal{T}^n X) = \mathbb{E}[G(X)] \quad \mathbb{P}\text{-a.s. and in mean.}$$

We deepen the study of empirical means investigating their normal fluctuations. Since X is a regenerative process with independent cycles [26], if G depends on exactly one binary variable, then the normal fluctuations of its empirical mean are described under the optimal hypothesis $\mathbb{E}[S_2^2] = \sum_{s\geq 1} s^2 p(s) < \infty$ by the central limit theorem for cumulative processes with a regenerative structure [26]. It is worth noting that in such case the empirical mean of G up to time t is a linear function of the number of renewals by t, so that even its large fluctuations can be completely characterized through well-established large deviation principles for discrete-time renewal-reward processes [38,39], which include the counting renewal process. To deal with functions G that involve more than one variable, and that cannot be tackled by the standard limit theory of regenerative processes with independent cycles, we resort to the theory of the central limit theorem for stationary sequences [40]. To begin with, we introduce the strong mixing coefficient α_t of the sequence X, which measures the memory of the past on future events t time steps later. According to [21], the *strong mixing coefficient α_t*, or *α-mixing coefficient*, of X is defined for each $t \geq 1$ by

$$\alpha_t := \sup_{m\geq 1}\ \sup_{\mathcal{A}\in\mathscr{F}_1^m}\ \sup_{\mathcal{B}\in\mathscr{F}_{m+t}^\infty} \left\{\Big|\mathbb{P}[\mathcal{A}\cap\mathcal{B}] - \mathbb{P}[\mathcal{A}]\,\mathbb{P}[\mathcal{B}]\Big|\right\}.$$

Here $\mathscr{F}_a^b$ is the σ-algebra generated by $X_a, \ldots, X_b$ for $1 \le a \le b \le \infty$. The sequence X is *strong mixing* in the sense of probability theory if $\lim_{t \uparrow \infty} \alpha_t = 0$. The following proposition provides an estimate of the α-mixing coefficient of our model. The proof is based on Lemma 3.1 and is proposed in Appendix K. Recall that $\rho_t := \mathrm{cov}[X_1, X_{t+1}]$.

Proposition 4.1 *For each $t \ge 1$*

$$\alpha_t \le 3\mu^2 \sum_{n \ge t} |\rho_{n+1} - \rho_n| + 4\mu^2 \sum_{n \ge t} n\, |\rho_{n+1} - 2\rho_n + \rho_{n-1}|.$$

Empirical means display normal fluctuations when the strong mixing coefficient decays reasonably fast, and precisely when $\sum_{t \ge 1} \alpha_t < \infty$. In fact, let us consider a bounded observable G and for each $n \ge 0$ let us set $Z_n := G(T^n X)$ for brevity. Due to boundedness of G Theorem 4.2 tells us that $\lim_{t \uparrow \infty}(1/t) \sum_{n=0}^{t-1} Z_n = \mathbb{E}[Z_0]$ $\mathbb{P}$-a.s. if p is aperiodic. The normal fluctuations of $(1/t) \sum_{n=0}^{t-1} Z_n$ around $\mathbb{E}[Z_0]$ are described as follows by Theorem 18.6.3 of [40], which states the central limit theorem for functionals of mixing sequences.

Theorem 4.3 *Let G be a bounded $\mathscr{B}$-measurable function on $\mathbb{N}_2^\infty$ and set $Z_n := G(T^n X)$ for $n \ge 0$. Assume that $\sum_{t \ge 1} \alpha_t < \infty$ and $\sum_{m \ge 1} \mathbb{E}[|Z_0 - \mathbb{E}[Z_0|\mathscr{F}_1^m]|] < \infty$. Then*

$$v := \mathrm{cov}[Z_0, Z_0] + 2 \sum_{n \ge 1} \mathrm{cov}[Z_0, Z_n]$$

is finite and non-negative, and provided that $v \ne 0$

$$\lim_{t \uparrow \infty} \mathbb{P}\left[\frac{1}{\sqrt{vt}} \sum_{n=0}^{t-1} (Z_n - \mathbb{E}[Z_0]) \le z \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}\zeta^2} d\zeta.$$

We point out that the hypotheses of the theorem about G are automatically satisfied when G depends only on a finite number of variables, since in such case $\mathbb{E}[Z_0|\mathscr{F}_1^m] = Z_0$ for all sufficiently large m. The following lemma based on Proposition 4.1 shows that the finiteness of the second moment of the waiting time distribution suffices for $\sum_{t \ge 1} \alpha_t < \infty$, thus implying the validity of the central limit theorem. The proof is provided in Appendix L.

Lemma 4.2 *The two following statements are equivalent:*

(i) p is aperiodic and $\sum_{s \ge 1} s^2 p(s) < \infty$;
(ii) $\sum_{t \ge 1} t |\rho_t - \rho_{t-1}| < \infty$.

Either of them implies $\sum_{t \ge 1} \alpha_t < \infty$ and $\sum_{s \ge 1} s^2 p(s) = \mu + 2\mu^3 \sum_{t \ge 0} \rho_t$.

We stress that a coupling argument can show, without the need of an explicit estimate of the α-mixing coefficient, that an aperiodic waiting time distribution satisfies $\sum_{s \ge 1} s^\gamma p(s) < \infty$ for a real number $\gamma > 1$ if and only if $\sum_{t \ge 1} t^{\gamma-2}\alpha_t < \infty$ (see [41], Theorem 6.1).

4.3 Empirical Analyses

Finally, we demonstrate the theory of empirical means through estimation of the waiting time distribution and of the autocovariance from data. The focus is on the possibility to identify their decay, that is on the possibility to estimate $p(s)$ for large s and ρ_τ for large τ. In order to avoid complications related to nonlinear functions of empirical means, we imagine that μ is known in advance. Importantly, we suppose that either (i) or (ii) of Lemma 4.2 hold, so
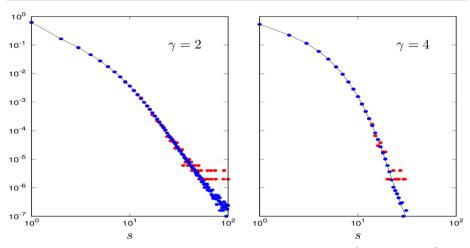
Fig. 3 Empirical estimates of $p(s)$ versus s with data sequences of length $t = 10^6$ (red) and $t = 10^8$ (blue) generated by the models defined by $c_t = 1/4 + (1/4)(1+t)^{-\gamma}$ for $t \geq 0$ with $\gamma = 2$ (left) and $\gamma = 4$ (right). Black curves are the theoretical limits $p(s)$ (Color figure online)

that if G is an observable that depends on a finite number of variables, then both the law of large numbers stated by Theorem 4.2 and the central limit theorem of Theorem 4.3 hold.

Fix an integer $s \geq 1$ and take

$$G(x) := \mu x_1 \prod_{k=2}^{s} (1 - x_k) x_{s+1}$$

for all $x \in \mathbb{N}_2^\infty$. Recalling that $Z_n := G(\mathcal{T}^n X)$ for $n \geq 0$, we have $\mathbb{E}[Z_0] = \mu \mathbb{P}[S_1 = 1, S_2 = s] = p(s)$, so that the empirical mean of G estimates the waiting time distribution at s. Simple applications of Proposition 3.1 show that $\text{cov}[Z_0, Z_n] = \mu p(s) - p^2(s)$ if $n = 0$, $\text{cov}[Z_0, Z_n] = -p^2(s)$ if $1 \leq n < s$, and $\text{cov}[Z_0, Z_n] = \mu^2 p^2(s) \rho_{n-s}$ if $n \geq s$. This way, the variance $v_s := \text{cov}[Z_0, Z_0] + 2 \sum_{n \geq 1} \text{cov}[Z_0, Z_n]$ introduced by Theorem 4.3 turns out to be

$$v_s = \mu p(s) + (1 - 2s) p^2(s) + 2\mu^2 p^2(s) \sum_{n \geq 0} \rho_n$$

$$= \mu p(s) - 2s p^2(s) + \mu^{-1} p^2(s) \sum_{n \geq 1} n^2 p(n).$$

We have made use of the identity $\sum_{s \geq 1} s^2 p(s) = \mu + 2\mu^3 \sum_{t \geq 0} \rho_t$ provided by Lemma 4.2 to get at the second equality. If a data sequence of length t is given, then we expect to be able to estimate $p(s)$ for values of s such that $\sqrt{v_s/t} \ll p(s)$. At large s, this means values of s such that $p(s) \gg \mu/t$ because $v_s \sim \mu p(s)$. Figure 3 shows results of estimation for two data sequences of length $t = 10^6$ and $t = 10^8$, respectively, when data are generated by the models of Fig. 1, which correspond to $c_t = 1/4 + (1/4)(1+t)^{-\gamma}$ for $t \geq 0$ with exponents $\gamma = 2$ and $\gamma = 4$. In such cases, the condition $p(s) \gg \mu/t$ for proper estimation reads $s \ll 1.56 \, t^{1/4}$ for $\gamma = 2$ and $s \ll 1.65 \, t^{1/6}$ for $\gamma = 4$. Figure 4 reports the same estimation for the models of Fig. 2 defined by $c_t = 1/4 + (1/4)e^{-t^\beta}$ for $t \geq 0$ with exponents $\beta = 1/2$ and $\beta = 1$. The condition $p(s) \gg \mu/t$ now becomes $s \ll \ln^2 t$ for $\beta = 1/2$ and $s \ll 2.63 \ln t$ for $\beta = 1$.
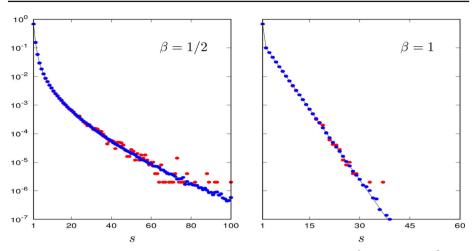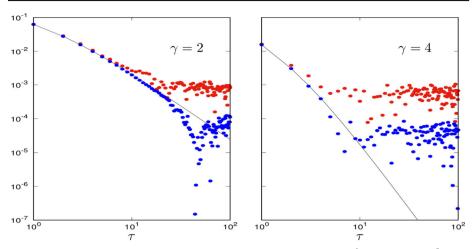
Fig. 4 Empirical estimates of $p(s)$ versus s with data sequences of length $t = 10^6$ (red) and $t = 10^8$ (blue) generated by the models defined by $c_t = 1/4 + (1/4)e^{-t^\beta}$ for $t \geq 0$ with $\beta = 1/2$ (left) and $\beta = 1$ (right). Black curves are the theoretical limits $p(s)$ (Color figure online)

Moving to the autocovariance, pick an integer $\tau \geq 0$ and consider the function

$$G(x) := x_1 x_{\tau+1} - 1/\mu^2 \qquad (4.2)$$

for all $x \in \mathbb{N}_2^\infty$. As $\mathbb{E}[Z_0] = \rho_\tau$, the empirical mean of F estimates ρ_τ. Once again, simple applications of Proposition 3.1 yield $\mathrm{cov}[Z_0, Z_n] = \mu^2 c_n^2 c_{\tau-n} - c_\tau^2$ if $0 \leq n < \tau$ and $\mathrm{cov}[Z_0, Z_n] = \mu^2 c_\tau^2 \rho_{n-\tau}$ if $n \geq \tau$. This way, the variance $v_\tau := \mathrm{cov}[Z_0, Z_0] + 2\sum_{n \geq 1} \mathrm{cov}[Z_0, Z_n]$ reads

$$
v_\tau = \begin{cases} \rho_0 + 2\sum_{n \geq 1} \rho_n & \text{if } \tau = 0; \\ c_\tau - c_\tau^2 + 2\sum_{n=1}^{\tau}(\mu^2 c_n^2 c_{\tau-n} - c_\tau^2) + 2\mu^2 c_\tau^2 \sum_{n \geq 0} \rho_n & \text{if } \tau \geq 1 \end{cases}
$$

$$
= \begin{cases} \mu^{-3} \sum_{n \geq 1} n^2 p(n) - \mu^{-1} & \text{if } \tau = 0; \\ 2\sum_{n=0}^{\tau}(\mu^2 c_n^2 c_{\tau-n} - c_\tau^2) + \mu^{-1} c_\tau^2 \sum_{n \geq 1} n^2 p(n) - c_\tau & \text{if } \tau \geq 1. \end{cases} \qquad (4.3)
$$

Since $\lim_{\tau \uparrow \infty} v_\tau = c_0^2(1 - 4c_0 + 3c_0^2) + 8c_0^2 \sum_{t \geq 0} \rho_t + 2\mu \sum_{t \geq 1} \rho_t^2 =: \sigma^2$ with $\sigma > 0$, we expect that ρ_τ can be estimated with a data sequence of length t when τ satisfies $\rho_\tau \gg \sigma/\sqrt{t}$. Figure 5 illustrates results of estimation for two data sequences of length $t = 10^6$ and $t = 10^8$, respectively, when data are generated by the models of Figs. 1 and 3. The condition $\rho_\tau \gg \sigma/\sqrt{t}$ explicitly is $\tau \ll 0.54\, t^{1/4}$ for $\gamma = 2$ and $\tau \ll 0.77\, t^{1/8}$ for $\gamma = 4$. Figure 6 provides the same results with reference to the models of Figs. 2 and 4. This time, the condition $\rho_\tau \gg \sigma/\sqrt{t}$ is $\tau \ll 0.25 \ln^2 t$ for $\beta = 1/2$ and $\tau \ll 0.5 \ln t$ for $\beta = 1$. The comparison between the waiting time distribution and the autocovariance shows that the former is easier to estimate than the latter.

A final remark is in order. If μ is not known in advance, then its inverse can be estimated through the empirical mean of the observable $G(x) := x_1$ for all $x \in \mathbb{N}_2^\infty$. Apart from an additive constant, this observable is (4.2) with $\tau = 0$. Thus, the variance of its empirical mean exactly is v_0 given by (4.3).

Fig. 5 Empirical estimates of ρ_τ versus τ with data sequences of length $t = 10^6$ (red) and $t = 10^8$ (blue) generated by the models defined by $c_t = 1/4 + (1/4)(1+t)^{-\gamma}$ for $t \geq 0$ with $\gamma = 2$ (left) and $\gamma = 4$ (right). Black curves are the theoretical limits ρ_τ (Color figure online)
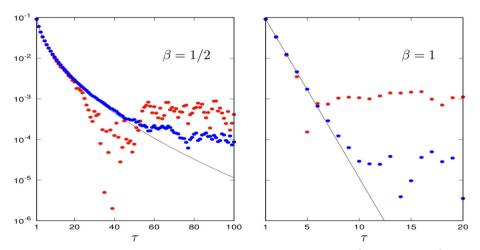

Fig. 6 Empirical estimates of ρ_τ versus τ with data sequences of length $t = 10^6$ (red) and $t = 10^8$ (blue) generated by the models defined by $c_t = 1/4 + (1/4)e^{-t^\beta}$ for $t \geq 0$ with $\beta = 1/2$ (left) and $\beta = 1$ (right). Black curves are the theoretical limits ρ_τ (Color figure online)

5 Conclusions

We have explored the use of renewal processes to generate binary sequences valued in $\{0, 1\}$, where the symbol 1 marks a renewal event. Focusing on stationary binary sequences corresponding to delayed renewal processes, we have demonstrated the ability of the model to account for subexponential autocovariances with special attention to polynomial and stretched-exponential decays. Our model performs at least as well as the algorithms proposed in [19] and [23,24] that seem to represent the state of the art, but at variance with them, generating a binary sequence with our model is a trivial task. Furthermore, our model is under full mathematical control, and this fact allowed us to build a mathematical theory of

its asymptotic properties. In fact, in addition to shedding light on the asymptotic behaviors of a large class of correlations and to demonstrating an asymptotic equipartition property, we have developed a theory for empirical means proving a law of large numbers and a central limit theorem. The latter describes the typical fluctuations of empirical means when the second moment of the waiting time distribution is finite. We leave for future work the study of typical fluctuations when the second moment of the waiting time distribution is infinite. In such case, empirical means are expected to be not in the Gaussian basin of attraction. We also leave for future work the investigation of their large fluctuations.

Data Availability No datasets were used to support this study.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

A Proof of Theorem 2.1

A monotone class argument [36] shows that the sequence X is stationary, namely that $\mathbb{P}[\{X_{t+1}\}_{t\geq 1} \in \mathcal{B}] = \mathbb{P}[\{X_t\}_{t\geq 1} \in \mathcal{B}]$ for all $\mathcal{B} \in \mathscr{B}$, if and only if $(X_2, \ldots, X_{t+1})$ is distributed as $(X_1, \ldots, X_t)$ for every $t \geq 1$. If the process X is stationary, then for any $s \geq 2$

$$
\begin{aligned}
\mathbb{P}[T_1 \geq s] = \mathbb{P}\big[X_1 = \cdots = X_{s-1} = 0\big] &= \mathbb{P}\big[X_2 = \cdots = X_s = 0\big] \\
&= \mathbb{P}\big[X_1 = 0, X_2 = \cdots = X_s = 0\big] \\
&\quad + \mathbb{P}\big[X_1 = 1, X_2 = \cdots = X_s = 0\big] \\
&= \mathbb{P}[T_1 \geq s+1] + \mathbb{P}[T_1 = 1, T_2 \geq s+1],
\end{aligned}
$$

which, due to the independence between $S_1 = T_1$ and $S_2 = T_2 - T_1$, is tantamount to

$$
\mathbb{P}[S_1 = s] = \mathbb{P}[S_1 = 1]\,\mathbb{P}[S_2 \geq s].
$$

This identity is trivial for $s = 1$ and is valid for all $s \geq 1$ as a consequence. In combination with the fact that $\sum_{s\geq 1}\mathbb{P}[S_1 = s] = 1$ and $\sum_{s\geq 1}\mathbb{P}[S_2 \geq s] = \mathbb{E}[S_2]$, it implies $\mathbb{P}[S_1 = 1] > 0$ and $1 = \mathbb{P}[S_1 = 1]\,\mathbb{E}[S_2]$. Thus, $\mathbb{E}[S_2] < \infty$ and $\mathbb{E}[S_2]\,\mathbb{P}[S_1 = s] = \mathbb{P}[S_2 \geq s]$ for every $s \geq 1$.

Assume now that $\mathbb{E}[S_2] < \infty$ and $\mathbb{E}[S_2]\,\mathbb{P}[S_1 = s] = \mathbb{P}[S_2 \geq s]$ for all $s \geq 1$. Let us prove that $(X_2, \ldots, X_{t+1})$ is distributed as $(X_1, \ldots, X_t)$ for any given $t \geq 1$. Pick binary numbers $x_1, \ldots, x_t$. By writing X_{k+1} as $\mathbb{1}_{\{k+1\in\{T_n\}_{n\geq 1}\}}$ and by observing that S_1 is independent of

$\{T_n - S_1\}_{n\geq 1}$ we find

$$
\mathbb{P}\big[X_2 = x_1, \ldots, X_{t+1} = x_t\big] = \mathbb{E}\left[\mathbb{1}_{\{S_1=1\}} \prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k+1\in\{T_n\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
+ \mathbb{E}\left[\mathbb{1}_{\{S_1>1\}} \prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k+1\in\{T_n\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
= \mathbb{E}\left[\mathbb{1}_{\{S_1=1\}} \prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1\}_{n\geq 2}\}}=x_k\right\}}\right]
$$

$$
+ \sum_{s\geq 1} \mathbb{E}\left[\mathbb{1}_{\{S_1=s+1\}} \prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1+s\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
= \mathbb{P}[S_1=1]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1\}_{n\geq 2}\}}=x_k\right\}}\right]
$$

$$
+ \sum_{s\geq 1} \mathbb{P}[S_1=s+1]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1+s\}_{n\geq 1}\}}=x_k\right\}}\right].
$$

At this point, we use the formula $\mathbb{P}[S_1=s+1] = \mathbb{P}[S_1=s] - \mathbb{P}[S_1=1]\,\mathbb{P}[S_2=s]$ valid for each s, which is an immediate consequence of the hypothesis $\mathbb{E}[S_2]\,\mathbb{P}[S_1=s] = \mathbb{P}[S_2\geq s]$. This way, by also noticing that $\{T_n - S_1\}_{n\geq 1}$ is distributed as $\{T_n - S_1 - S_2\}_{n\geq 2}$, we get

$$
\mathbb{P}\big[X_2 = x_1, \ldots, X_{t+1} = x_t\big]
$$

$$
= \mathbb{P}[S_1=1]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1\}_{n\geq 2}\}}=x_k\right\}}\right]
$$

$$
+ \sum_{s\geq 1} \mathbb{P}[S_1=s]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1+s\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
- \mathbb{P}[S_1=1]\sum_{s\geq 1} \mathbb{P}[S_2=s]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1-S_2+s\}_{n\geq 2}\}}=x_k\right\}}\right].
$$

The third term of the r.h.s. cancels with the first one since the independence between S_2 and $\{T_n - S_1 - S_2 + s\}_{n\geq 2}$ shows that the latter is an expanded version of the former. Then

$$
\mathbb{P}\big[X_2 = x_1, \ldots, X_{t+1} = x_t\big] = \sum_{s\geq 1} \mathbb{P}[S_1=s]\,\mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n-S_1+s\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
= \mathbb{E}\left[\prod_{k=1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k\in\{T_n\}_{n\geq 1}\}}=x_k\right\}}\right]
$$

$$
= \mathbb{P}\big[X_1 = x_1, \ldots, X_t = x_t\big].
$$

B Proof of Proposition 2.1

Pick an integer $t \geq 1$ and arbitrary numbers $x_1, x_2, \ldots, x_t$ in $\mathbb{N}_2$. If $x_1 = x_2 = \cdots = x_t = 0$, then

$$\pi_t(x_1, x_2, \ldots, x_t) := \mathbb{P}\big[X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t\big]$$
$$= \mathbb{P}\big[T_1 > t\big] = \mathbb{P}\big[S_1 > t\big] = \frac{1}{\mu} \sum_{s \geq t} Q(s).$$

If $x_1, x_2, \ldots, x_t$ account for exactly $N \geq 1$ ones in positions $1 \leq i_1 < i_2 < \cdots < i_N \leq t$, then

$$\pi_t(x_1, x_2, \ldots, x_t) = \mathbb{P}\big[T_1 = i_1, T_2 = i_2, \ldots, T_N = i_N, T_{N+1} > t\big]$$
$$= \mathbb{P}\big[S_1 = i_1, S_2 = i_2 - i_1, \ldots, S_N = i_N - i_{N-1}, S_{N+1} > t - i_N\big]$$
$$= \mathbb{P}\big[S_1 = i_1\big] \prod_{n=2}^{N} \mathbb{P}\big[S_n = i_n - i_{n-1}\big] \mathbb{P}\big[S_{N+1} > t - i_N\big]$$
$$= \frac{1}{\mu} Q(i_1 - 1) \prod_{n=2}^{N} p(i_n - i_{n-1}) Q(t - i_N).$$

These formulas together with the convention $0^0 := 1$ show that

$$\pi_t(x_1, x_2, \ldots, x_t) = \frac{1}{\mu}\left[\sum_{s \geq t} Q(s)\right]^{\prod_{n=1}^{t}(1-x_n)} \prod_{i=1}^{t} \big[Q(i-1)\big]^{\prod_{n=1}^{i-1}(1-x_n)x_i}$$
$$\cdot \prod_{i=1}^{t-1} \prod_{j=i+1}^{t} \big[p(j-i)\big]^{x_i \prod_{n=i+1}^{j-1}(1-x_n)x_j} \prod_{j=1}^{t} \big[Q(t-j)\big]^{x_j \prod_{n=j+1}^{t}(1-x_n)}.$$

Finally, by changing i with $t - j + 1$ and j with $t - i + 1$ in the above products we realize that

$$\pi_t(x_1, x_2, \ldots, x_t) = \pi_t(x_t, \ldots, x_2, x_1).$$

C Proof of Corollary 2.1

Thanks to the independence between S_1 and S_2, for all $t \geq 1$ we have

$$\mathbb{P}\big[X_1 = 1, X_2 = 0, \ldots, X_{t+1} = 0\big] = \mathbb{P}\big[T_1 = 1, T_2 > t + 1\big] = \mathbb{P}\big[S_1 = 1, S_2 > t\big] = \frac{Q(t)}{\mu}.$$

Thus, if X is a sequence of i.i.d. binary random variables with mean $1/\mu$, then for each t

$$Q(t) = \left(1 - \frac{1}{\mu}\right)^t.$$

Setting $\lambda := 1 - 1/\mu \in [0, 1)$, this identity shows that $p(s + 1) = \lambda^s p(1)$ for all $s \geq 1$. On the contrary, if $p(s + 1) = \lambda^s p(1)$ for all $s \geq 1$ with some real number $\lambda \in [0, 1)$, then Proposition 2.1 demonstrates that X is a sequence of i.i.d. random variables after some simple algebra.

Let us move to discuss Markovianity of order $M \geq 1$. The sequence X is a Markov chain of order $M \geq 1$ if for all $t > M$ and $x_1, \ldots, x_{t+1}$ such that $\mathbb{P}[X_1 = x_1, \ldots, X_t = x_t] > 0$

$$\mathbb{P}\big[X_{t+1} = x_{t+1}\big|X_t = x_t, \ldots, X_1 = x_1\big]$$
$$= \mathbb{P}\big[X_{t+1} = x_{t+1}\big|X_t = x_t, \ldots, X_{t-M+1} = x_{t-M+1}\big].$$

The constraint $\mathbb{P}[X_1 = x_1, \ldots, X_t = x_t] > 0$ can be dropped by multiplying both sides by $\mathbb{P}[X_1 = x_1, \ldots, X_t = x_t]$ and $\mathbb{P}[X_1 = x_1, \ldots, X_M = x_M]$. In fact, the above condition is tantamount to ask the for all $t > M$ and $x_1, \ldots, x_{t+1}$

$$\mathbb{P}\big[X_1 = x_1, \ldots, X_{t+1} = x_{t+1}\big] \, \mathbb{P}\big[X_{t-M+1} = x_{t-M+1}, \ldots, X_t = x_t\big]$$
$$= \mathbb{P}\big[X_1 = x_1, \ldots, X_t = x_t\big] \, \mathbb{P}\big[X_{t-M+1} = x_{t-M+1}, \ldots, X_{t+1} = x_{t+1}\big],$$

which, due to stationarity, reads

$$\mathbb{P}\big[X_1 = x_1, \ldots, X_{t+1} = x_{t+1}\big] \, \mathbb{P}\big[X_1 = x_{t-M+1}, \ldots, X_M = x_t\big]$$
$$= \mathbb{P}\big[X_1 = x_1, \ldots, X_t = x_t\big] \, \mathbb{P}\big[X_1 = x_{t-M+1}, \ldots, X_{M+1} = x_{t+1}\big]. \tag{C.1}$$

By making the choice $x_1 = 1$ and $x_2 = \cdots = x_{t+1} = 0$ in (C.1) we get for each $t > M$

$$\mathbb{P}\big[S_1 = 1, S_2 > t\big] \, \mathbb{P}\big[S_1 > M\big] = \mathbb{P}\big[S_1 = 1, S_2 > t - 1\big] \, \mathbb{P}\big[S_1 > M + 1\big],$$

which, by appealing to the independence between S_1 and S_2 once more, reduces to

$$Q(t) \, Q(M) = Q(t - 1) \, Q(M + 1). \tag{C.2}$$

This is a necessary condition for Markovianity of order $M \geq 1$.

In order to find a sufficient condition, we plug in (C.1) the explicit expression of finite-dimensional distributions given by Proposition 2.1. By equating the factors that the two sides of the equation do not have in common we reach the condition

$$p(t) \sum_{s \geq M} Q(s) = Q(t - 1) \, Q(M) \tag{C.3}$$

to be satisfied for every $t > M$.

Condition (C.2) imposes that either $Q(M) = 0$ or $p(M + s + 1) = \lambda^s \, p(M + 1)$ for all $s \geq 1$ with $\lambda := Q(M + 1)/Q(M) \in [0, 1)$. Since $Q(M) = 0$ implies $p(M + s + 1) = 0$ for each $s \geq 0$, in any case we can state that (C.2) entails that there exists a real number $\lambda \in [0, 1)$ such that $p(M + s + 1) = \lambda^s \, p(M + 1)$ for all $s \geq 1$. A waiting time distribution that fulfills such relation with some $\lambda \in [0, 1)$ solves (C.3), as one can easily verify.

D Proof of Proposition 3.1

Fix positive integers $s \leq t$ and arbitrary binary numbers $x_1, \ldots, x_t$. The proposition is trivial if $s = 1$, $s = t$, or $1 < s < t$ and $x_s = 0$. Then, assume $1 < s < t$ and $x_s = 1$. Since the condition $X_s = 1$ implies that $s \in \{T_m\}_{m \geq 1}$ and since $T_1 < T_2 < \cdots$, we can write

$$\mathbb{P}\big[X_1 = x_1, \ldots, X_t = x_t\big] = \mathbb{E}\Bigg[\prod_{k=1}^{s-1} \mathbb{1}_{\big\{\mathbb{1}_{\{k \in \{T_n\}_{n \geq 1}\}} = x_k\big\}} \mathbb{1}_{\{s \in \{T_m\}_{m \geq 1}\}} \prod_{k=s+1}^{t} \mathbb{1}_{\big\{\mathbb{1}_{\{k \in \{T_n\}_{n \geq 1}\}} = x_k\big\}}\Bigg]$$

$$= \sum_{m \geq 1} \mathbb{E}\Bigg[\prod_{k=1}^{s-1} \mathbb{1}_{\big\{\mathbb{1}_{\{k \in \{T_n\}_{n=1}^m\}} = x_k\big\}} \mathbb{1}_{\{T_m = s\}}$$

$$\times \prod_{k=s+1}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k-s \in \{T_n - T_m\}_{n \geq m}\}} = x_k\right\}}\Bigg].$$

For each $m \geq 1$, the sequence $\{T_n - T_m\}_{n \geq m}$ is independent of $\{T_n\}_{n=1}^{m}$ and distributed as $\{T_n - T_1\}_{n \geq 1}$, so that

$$\mathbb{P}\big[X_1 = x_1, \ldots, X_t = x_t\big]$$

$$= \sum_{m \geq 1} \mathbb{E}\Bigg[\prod_{k=1}^{s-1} \mathbb{1}_{\left\{\mathbb{1}_{\{k \in \{T_n\}_{n=1}^{m}\}} = x_k\right\}} \mathbb{1}_{\{T_m = s\}}\Bigg] \mathbb{E}\Bigg[\prod_{k=s}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k-s \in \{T_n - T_1\}_{n \geq 1}\}} = x_k\right\}}\Bigg]$$

$$= \mathbb{E}\Bigg[\prod_{k=1}^{s-1} \mathbb{1}_{\left\{\mathbb{1}_{\{k \in \{T_n\}_{n \geq 1}\}} = x_k\right\}} \mathbb{1}_{\{s \in \{T_m\}_{m \geq 1}\}}\Bigg] \mathbb{E}\Bigg[\prod_{k=s}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k-s \in \{T_n - T_1\}_{n \geq 1}\}} = x_k\right\}}\Bigg]$$

$$= \mathbb{P}\big[X_1 = x_1, \ldots, X_s = x_s\big] \mathbb{E}\Bigg[\prod_{k=s}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k-s \in \{T_n - T_1\}_{n \geq 1}\}} = x_k\right\}}\Bigg].$$

This identity proves the proposition since by taking the sum over $x_1, \ldots, x_{s-1}$ we realize that

$$\mathbb{E}\Bigg[\prod_{k=s}^{t} \mathbb{1}_{\left\{\mathbb{1}_{\{k-s \in \{T_n - T_1\}_{n \geq 1}\}} = x_k\right\}}\Bigg] = \mathbb{P}\big[X_s = x_s, \ldots, X_t = x_t \big| X_s = 1\big].$$

E Proof of Proposition 3.2

Pick $t \geq 1$. The identity $1 = \sum_{s=1}^{t} X_{t-s+1} \prod_{k=t-s+2}^{t}(1 - X_k) + \prod_{k=1}^{t}(1 - X_k)$ results in the formula $X_1 = \sum_{s=1}^{t} X_1 X_{t-s+1} \prod_{k=t-s+2}^{t}(1 - X_k)$, which allows us to write down

$$c_t := \mathbb{E}\big[X_1 X_{t+1}\big] = \sum_{s=1}^{t} \mathbb{E}\Bigg[X_1 X_{t-s+1} \prod_{k=t-s+2}^{t}(1 - X_k) X_{t+1}\Bigg]$$

$$= \sum_{s=1}^{t} \mathbb{E}\Bigg[X_1 \prod_{k=t-s+2}^{t}(1 - X_k) X_{t+1} \Big| X_{t-s+1} = 1\Bigg] \mathbb{E}[X_1].$$

This way, Proposition 3.1 shows that

$$c_t = \sum_{s=1}^{t} \mathbb{E}\big[X_1 \big| X_{t-s+1} = 1\big] \mathbb{E}\Bigg[\prod_{k=t-s+2}^{t}(1 - X_k) X_{t+1} \Big| X_{t-s+1} = 1\Bigg] \mathbb{E}[X_1]$$

$$= \sum_{s=1}^{t} \frac{\mathbb{E}[X_{t-s+1} \prod_{k=t-s+2}^{t}(1 - X_k) X_{t+1}]}{\mathbb{E}[X_1]} c_{t-s}.$$

Finally, stationarity yields

$$c_t = \sum_{s=1}^{t} \frac{\mathbb{E}[X_1 \prod_{k=2}^{s}(1 - X_k) X_{s+1}]}{\mathbb{E}[X_1]} c_{t-s}$$

$$= \sum_{s=1}^{t} \frac{\mathbb{P}[T_1 = 1, T_2 = s + 1]}{\mathbb{P}[T_1 = 1]} c_{t-s} = \sum_{s=1}^{t} \mathbb{P}[S_2 = s] c_{t-s} = \sum_{s=1}^{t} p(s) c_{t-s}.$$

F Proof of Lemma 3.1

Set for brevity $Y := f(X_1, \ldots, X_m)$, $\Phi_i := X_{m-i+1} \prod_{k=m-i+2}^{m}(1 - X_k)$ for $1 \le i \le m$, and $\Psi_j(t) := \prod_{k=1}^{j-1}(1 - X_{m+t+k-1})X_{m+t+j-1}$ for $j \ge 1$ and $t \ge 1$. We are going to show that for all $t \ge 1$

$$\text{cov}\big[Y, Z_{m+t}\big] = \sum_{i=1}^{m}\sum_{j=1}^{n} \text{cov}\big[\Phi_i, \Psi_j(t)\big] \frac{\mathbb{E}[f(X_m, \ldots, X_1)\mathbb{1}_{\{S_1=i\}}]}{\mathbb{P}[S_1 = i]} \frac{\mathbb{E}[\mathbb{1}_{\{S_1=j\}}Z_1]}{\mathbb{P}[S_1 = j]} \quad \text{(F.1)}$$

and, for every $i \ge 1$ and $j \ge 1$,

$$\text{cov}\big[\Phi_i, \Psi_j(t)\big] = C_{i,j}(t). \quad \text{(F.2)}$$

We verify (F.1) at first. Pick $t \ge 1$. The identities $1 = \sum_{i=1}^{m}\Phi_i + \prod_{k=1}^{m}(1 - X_k)$ and $1 = \sum_{j=1}^{n}\Psi_j(t) + \prod_{k=0}^{n-1}(1 - X_{m+t+k})$ give $Y = \sum_{i=1}^{m} Y\Phi_i$ and $Z_{m+t} = \sum_{j=1}^{n}\Psi_j(t)Z_{m+t}$ as $f(0, \ldots, 0) = 0$ and $g(0, \ldots, 0) = 0$ by hypothesis. Then, we can write down the formula

$$\text{cov}\big[Y, Z_{m+t}\big] = \sum_{i=1}^{m}\sum_{j=1}^{n}\left\{ \mathbb{E}\big[Y\Phi_i\Psi_j(t)Z_{m+t}\big] - \mathbb{E}\big[Y\Phi_i\big]\mathbb{E}\big[\Psi_j(t)Z_{m+t}\big]\right\}. \quad \text{(F.3)}$$

Let us manipulate $\mathbb{E}[Y\Phi_i\Psi_j(t)Z_{m+t}]$ for two given indices $i \le m$ and $j \le n$. The condition $\Phi_i \ne 0$ implies $X_{m-i+1} = 1$ and $X_{m-i+2} = \cdots = X_m = 0$. This way, Proposition 3.1 yields

$$\frac{\mathbb{E}[Y\Phi_i]}{\mathbb{E}[X_{m-i+1}]} = \mathbb{E}\Big[f(X_1, \ldots, X_{m-i+1}, 0, \ldots, 0)\Phi_i \,\Big|\, X_{m-i+1} = 1\Big]$$

$$= \mathbb{E}\Big[f(X_1, \ldots, X_{m-i+1}, 0, \ldots, 0)\,\Big|\, X_{m-i+1} = 1\Big]\mathbb{E}\Big[\Phi_i \,\Big|\, X_{m-i+1} = 1\Big]$$

$$= \frac{\mathbb{E}[f(X_1, \ldots, X_{m-i+1}, 0, \ldots, 0)|X_{m-i+1} = 1]\mathbb{E}[\Phi_i]}{\mathbb{E}[X_{m-i+1}]},$$

that is

$$\mathbb{E}\big[Y\Phi_i\big] = \mathbb{E}\big[f(X_1, \ldots, X_{m-i+1}, 0, \ldots, 0)\big|X_{m-i+1} = 1\big]\mathbb{E}\big[\Phi_i\big].$$

Similarly, as the random quantity $\Psi_j(t)Z_{m+t}$ is a function of $X_{m+t}, \ldots, X_{m+t+n-1}$ only, we have

$$\mathbb{E}\big[Y\Phi_i\Psi_j(t)Z_{m+t}\big] = \mathbb{E}\big[f(X_1, \ldots, X_{m-i+1}, 0, \ldots, 0)\big|X_{m-i+1} = 1\big]\mathbb{E}\big[\Phi_i\Psi_j(t)Z_{m+t}\big].$$

By continuing with these arguments, since the condition $\Psi_j(t) \ne 0$ entails $X_{m+t} = \cdots = X_{m+t+j-2} = 0$ and $X_{m+t+j-1} = 1$, we realize that

$$\mathbb{E}\big[\Psi_j(t)Z_{m+t}\big] = \mathbb{E}\big[\Psi_j(t)\big]\mathbb{E}\big[g(0, \ldots, 0, X_{m+t+j-1}, \ldots, X_{m+t+n-1})\big|X_{m+t+j-1} = 1\big]$$

and

$$\mathbb{E}\big[\Phi_i\Psi_j(t)Z_{m+t}\big]$$
$$= \mathbb{E}\big[\Phi_i\Psi_j(t)\big]\mathbb{E}\big[g(0, \ldots, 0, X_{m+t+j-1}, \ldots, X_{m+t+n-1})\big|X_{m+t+j-1} = 1\big].$$

These four relations show that

$$\mathbb{E}\big[Y\Phi_i\Psi_j(t)Z_{m+t}\big] = \frac{\mathbb{E}[Y\Phi_i]\mathbb{E}[\Phi_i\Psi_j(t)]\mathbb{E}[\Psi_j(t)Z_{m+t}]}{\mathbb{E}[\Phi_i]\mathbb{E}[\Psi_j(t)]}. \quad \text{(F.4)}$$

On the other hand, reversibility of X gives

$$\mathbb{E}[Y\Phi_i] = \mathbb{E}\left[f(X_m, \ldots, X_1)\prod_{k=1}^{i-1}(1 - X_k)X_i\right] = \mathbb{E}\left[f(X_m, \ldots, X_1)\mathbb{1}_{\{S_1=i\}}\right] \quad \text{(F.5)}$$

and

$$\mathbb{E}[\Phi_i] = \mathbb{P}[S_1 = i]. \tag{F.6}$$

Stationarity of X implies

$$\mathbb{E}\left[\Psi_j(t)Z_{m+t}\right] = \mathbb{E}\left[\prod_{k=1}^{j-1}(1 - X_k)X_j Z_1\right] = \mathbb{E}\left[\mathbb{1}_{\{S_1=j\}}Z_1\right] \tag{F.7}$$

and

$$\mathbb{E}[\Psi_j(t)] = \mathbb{P}[S_1 = j]. \tag{F.8}$$

Formula (F.3) in combination with (F.4), (F.5), (F.6), (F.7), and (F.8) results in (F.1).

Let us move to (F.2), which we prove by induction with respect to t. To begin with, let us verify that $\mathrm{cov}[\Phi_i, \Psi_j(1)] = C_{i,j}(1)$ for each $i \geq 1$ and $j \geq 1$. By recalling (F.6) and (F.8) and by invoking stationarity of X we find

$$\begin{aligned}
\mathrm{cov}\left[\Phi_i, \Psi_j(1)\right] &= \mathbb{E}\left[X_{m-i+1}\prod_{k=m-i+2}^{m+j-1}(1 - X_k)X_{m+j}\right] - \mathbb{E}[\Phi_i]\,\mathbb{E}[\Psi_j(1)]\\
&= \mathbb{E}\left[X_1\prod_{k=2}^{i+j-1}(1 - X_k)X_{i+j}\right] - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]\\
&= \mathbb{P}[S_1 = 1, S_2 = i + j - 1] - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]\\
&= \mathbb{P}[S_1 = 1]\,p(i + j - 1) - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]. \tag{F.9}
\end{aligned}$$

On the other hand, by definition we have

$$\begin{aligned}
C_{i,j}(1) &= \sum_{u=1}^{i}\sum_{v=1}^{j} p(i - u)\,\rho_{u+v-1}\,p(j - v)\\
&= \sum_{u=1}^{i}\sum_{v=1}^{j} p(i - u)\,c_{u+v-1}\,p(j - v) - \sum_{u=1}^{i}\sum_{v=1}^{j} p(i - u)\,c_0^2\,p(j - v)\\
&= \sum_{u=0}^{i-1} p(u)\sum_{v=0}^{j-1} c_{i+j-u-1-v}\,p(v) - \frac{1}{\mu}\sum_{u=0}^{i-1} p(u)\,\frac{1}{\mu}\sum_{v=0}^{j-1} p(v).
\end{aligned}$$

Proposition 3.2 tells us that $\sum_{v=1}^{i+j-u-1} c_{i+j-u-1-v}\,p(v) = c_{i+j-u-1}$ if $u \leq i - 1$. Then, by recalling that $p(0) := -1$, we realize that $\sum_{v=0}^{j-1} c_{i+j-u-1-v}\,p(v) = -\sum_{v=j}^{i+j-u-1} c_{i+j-u-1-v}\,p(v)$ for $u \leq i - 1$. The fact that $p(0) := -1$ also gives $\sum_{u=0}^{i-1} p(u) = -\mathbb{P}[S_2 \geq i]$ and $\sum_{v=0}^{j-1} p(v) = -\mathbb{P}[S_2 \geq j]$. Since $(1/\mu)\mathbb{P}[S_2 \geq i] = \mathbb{P}[S_1 = i]$ and $(1/\mu)\mathbb{P}[S_2 \geq j] = \mathbb{P}[S_1 = j]$ by Theorem 2.1, we obtain

$$C_{i,j}(1) = -\sum_{u=0}^{i-1} p(u) \sum_{v=j}^{i+j-u-1} c_{i+j-u-1-v}\, p(v) - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]$$

$$= -\sum_{v=j}^{i+j-1} p(v) \sum_{u=0}^{i+j-v-1} c_{i+j-v-1-u}\, p(u) - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j].$$

Another application of Proposition 3.2 yields $\sum_{u=0}^{i+j-v-1} c_{i+j-v-1-u}\, p(u) = 0$ if $v < i + j - 1$ and $\sum_{u=0}^{i+j-v-1} c_{i+j-v-1-u}\, p(u) = -c_0$ if $v = i + j - 1$. Then

$$C_{i,j}(1) = p(i + j - 1)c_0 - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]$$
$$= \mathbb{P}[S_1 = 1]\, p(i + j - 1) - \mathbb{P}[S_1 = i]\,\mathbb{P}[S_1 = j]. \tag{F.10}$$

Formulas (F.9) and (F.10) demonstrate that $\mathrm{cov}[\Phi_i, \Psi_j(1)] = C_{i,j}(1)$ for each $i \geq 1$ and $j \geq 1$.

To conclude, let us show that if $\mathrm{cov}[\Phi_i, \Psi_j(t)] = C_{i,j}(t)$ for all positive indices i and j and some $t \geq 1$, then $\mathrm{cov}[\Phi_i, \Psi_j(t + 1)] = C_{i,j}(t + 1)$. Fix $i \geq 1$ and $j \geq 1$. By using the identity $\Psi_j(t + 1) = \Psi_{j+1}(t) + X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}$ and the induction hypothesis we get

$$\mathrm{cov}\big[\Phi_i, \Psi_j(t + 1)\big] = C_{i,j+1}(t) + \mathrm{cov}\left[\Phi_i, X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}\right].$$

On the other hand, Proposition 3.1 allows the factorization

$$\mathbb{E}\left[\Phi_i X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}\right]$$
$$= \frac{\mathbb{E}[\Phi_i X_{m+t}]\,\mathbb{E}[X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}]}{\mathbb{E}[X_{m+t}]},$$

which results in

$$\mathrm{cov}\left[\Phi_i, X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}\right]$$
$$= \frac{\mathbb{E}[\Phi_i X_{m+t}]\,\mathbb{E}[X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}]}{\mathbb{E}[X_{m+t}]}$$
$$- \mathbb{E}[\Phi_i]\left[X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}\right]$$
$$= \frac{\mathrm{cov}[\Phi_i, X_{m+t}]\,\mathbb{E}[X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}]}{\mathbb{E}[X_{m+t}]}.$$

Since $X_{m+t} = \Psi_1(t)$, the induction hypothesis and stationarity of X give

$$\mathrm{cov}\left[\Phi_i, X_{m+t} \prod_{k=1}^{j-1}(1 - X_{m+t+k})X_{m+t+j}\right] = \frac{C_{i,1}(t)\,\mathbb{E}[X_1 \prod_{k=2}^{j}(1 - X_k)X_{j+1}]}{\mathbb{E}[X_1]}$$
$$= \frac{C_{i,1}(t)\,\mathbb{P}[S_1 = 1, S_2 = j]}{\mathbb{P}[S_1 = 1]} = C_{i,1}(t)\, p(j).$$

It follows that

$$\mathrm{cov}\big[\Phi_i, \Psi_j(t+1)\big] = C_{i,j+1}(t) + C_{i,1}(t)\,p(j)$$

$$= \sum_{u=1}^{i}\sum_{v=1}^{j+1} p(i-u)\,\rho_{u+t+v-2}\,p(j-v+1)$$

$$- \sum_{u=1}^{i} p(i-u)\,\rho_{u+t-1}\,p(j)$$

$$= \sum_{u=1}^{i}\sum_{v=1}^{j} p(i-u)\,\rho_{u+t+1+v-2}\,p(j-v) = C_{i,j}(t+1).$$

G Proof of Corollary 3.2

Set $\sigma(a) := \sup_{z\geq a}\{|z\ell'(z)/\ell(z)|\}$ and observe that $\lim_{a\uparrow\infty}\sigma(a) = 0$ by hypothesis. The inequality $|\ln\ell(b)-\ln\ell(a)| \leq \int_a^b |\ell'(z)/\ell(z)|dz \leq \sigma(a)(\ln b-\ln a)$ gives for all sufficiently large $a \leq b$ the bound

$$\left(\frac{a}{b}\right)^{\sigma(a)} \leq \frac{\ell(b)}{\ell(a)} \leq \left(\frac{b}{a}\right)^{\sigma(a)}. \tag{G.1}$$

This bound shows that ℓ is slowly varying since $\lim_{a\uparrow\infty}\sigma(a) = 0$. It also shows that $\beta z^{\beta-1}\ell(z)/2 \leq (z+1)^\beta \ell(z+1) - z^\beta\ell(z) \leq 2\beta z^{\beta-1}\ell(z)$ for each z large enough, so that $\lim_{z\uparrow\infty}[(z+1)^\beta\ell(z+1) - z^\beta\ell(z)] = 0$ due to the fact that ℓ is dominated by polynomials according to Proposition 1.3.6 of [34]. Then, $Q(t+1) \sim e^{-(t+1)^\beta\ell(t+1)} \sim e^{-t^\beta\ell(t)} \sim Q(t)$.

Let h be the continuously differentiable real function on a neighborhood of infinity that maps z in $h(z) := z^{\beta-1}\ell(z)$. We now demonstrate that $-z^2 h'(z)$ is increasing to infinity with respect to z and that $\int_\tau^\infty e^{\frac{1}{2}z^2 h'(z)}dz < \infty$ for some $\tau > 0$. It follows that $\{Q(t)\}_{t\geq 0} \in \mathscr{S}$ by the condition for subexponentiality given in [32], so that Theorem 3.1 ensures us that $\rho_t \sim (1/\mu^3)\sum_{n>t} Q(n)$. As $\lim_{z\uparrow\infty} z\ell'(z)/\ell(z) = 0$ by hypothesis, the formula

$$\frac{z^2 h'(z)}{z^\beta \ell(z)} = \beta - 1 + \frac{z\ell'(z)}{\ell(z)}$$

yields

$$\lim_{z\uparrow\infty}\frac{z^2 h'(z)}{z^\beta\ell(z)} = \beta - 1 < 0.$$

Since ℓ is dominated by polynomials, this limit entails $\lim_{z\uparrow\infty} -z^2 h'(z) = \infty$ and at the same time shows that $\int_\tau^\infty e^{\frac{1}{2}z^2 h'(z)}dz < \infty$ for some τ large enough. It remains to verify that $-z^2 h'(z)$ is increasing with respect to z. The condition $\lim_{z\uparrow\infty} z\ell'(z)/\ell(z) = 0$ implies $\lim_{z\uparrow\infty} z^2\ell''(z)/\ell(z) = 0$ by L'Hôpital's rule as $\lim_{z\uparrow\infty} z^2\ell''(z)/\ell(z)$ is assumed to exist. Then, the formula

$$\frac{[-z^2 h'(z)]'}{z^{\beta-1}\ell(z)} = \beta(1-\beta) - 2\beta\frac{z\ell'(z)}{\ell(z)} - \frac{z^2\ell''(z)}{\ell(z)}$$

results in

$$\lim_{z\uparrow\infty} \frac{[-z^2 h'(z)]'}{z^{\beta-1}\ell(z)} = \beta(1-\beta) > 0.$$

This suffices to state that $-z^2 h'(z)$ is increasing with respect to z.

To conclude, let us prove that

$$\sum_{n>t} Q(n) \sim \frac{t^{1-\beta}}{\beta\ell(t)} e^{-t^\beta\ell(t)}. \tag{G.2}$$

Pick $\epsilon \in (0, \beta \wedge 1 - \beta)$ and recall that $\lim_{a\uparrow\infty}\sigma(a) = 0$ by hypothesis. The fact that $Q(t) \sim e^{-t^\beta\ell(t)}$ and (G.1) imply that there exists $t_o > 0$ such that $(1-\epsilon)e^{-n^\beta\ell(n)} \le Q(n) \le (1+\epsilon)e^{-n^\beta\ell(n)}$ and $t^\epsilon n^{\beta-\epsilon}\ell(t) \le n^\beta\ell(n) \le n^{\beta+\epsilon}(t+1)^{-\epsilon}\ell(t+1)$ for $n > t > t_o$. This way, for every $t > t_o$ we get

$$\sum_{n>t} Q(n) \le (1+\epsilon)\sum_{n>t} e^{-n^\beta\ell(n)} \le (1+\epsilon)\sum_{n>t} e^{-n^{\beta-\epsilon}t^\epsilon\ell(t)}$$

$$\le (1+\epsilon)\int_{z>t} e^{-z^{\beta-\epsilon}t^\epsilon\ell(t)}dz$$

$$= \frac{1+\epsilon}{\beta-\epsilon}\left[\frac{1}{t^\epsilon\ell(t)}\right]^{\frac{1}{\beta-\epsilon}}\int_{\zeta>t^\beta\ell(t)} \zeta^{\frac{1-\beta+\epsilon}{\beta-\epsilon}} e^{-\zeta}d\zeta$$

and

$$\sum_{n>t} Q(n) \ge (1-\epsilon)\sum_{n>t} e^{-n^\beta\ell(n)} \ge (1-\epsilon)\sum_{n>t} e^{-n^{\beta+\epsilon}(t+1)^{-\epsilon}\ell(t+1)}$$

$$\ge (1-\epsilon)\int_{z>t+1} e^{-z^{\beta+\epsilon}(t+1)^{-\epsilon}\ell(t+1)}dz$$

$$= \frac{1-\epsilon}{\beta+\epsilon}\left[\frac{(t+1)^\epsilon}{\ell(t+1)}\right]^{\frac{1}{\beta+\epsilon}}\int_{\zeta>(t+1)^\beta\ell(t+1)} \zeta^{\frac{1-\beta-\epsilon}{\beta+\epsilon}} e^{-\zeta}d\zeta.$$

An integration by part shows that $a^\gamma e^{-a} \le \int_{\zeta>a} \zeta^\gamma e^{-\zeta}d\zeta \le a^\gamma e^{-a}/(1-\gamma/a)$ for every real numbers $\gamma > 0$ and $a > \gamma$. Then

$$\sum_{n>t} Q(n) \le \frac{1+\epsilon}{\beta-\epsilon}\frac{t^{1-\beta}}{\ell(t)}\frac{e^{-t^\beta\ell(t)}}{1-\frac{1-\beta+\epsilon}{\beta-\epsilon}\frac{1}{t^\beta\ell(t)}}$$

and

$$\sum_{n>t} Q(n) \ge \frac{1-\epsilon}{\beta+\epsilon}\frac{(t+1)^{1-\beta}}{\ell(t+1)} e^{-(t+1)^\beta\ell(t+1)}.$$

By recalling that ℓ is dominated by polynomials and that $\lim_{z\uparrow\infty}[(z+1)^\beta\ell(z+1)-z^\beta\ell(z)] = 0$ and by observing that $\ell(t+1) \sim \ell(t)$ by the bound $(1+1/t)^{-\sigma(t)} \le \ell(t+1)/\ell(t) \le (1+1/t)^{\sigma(t)}$ valid for sufficiently large t, we find

$$\frac{1-\epsilon}{\beta+\epsilon} \le \liminf_{t\uparrow\infty} \ell(t)t^{\beta-1}e^{t^\beta\ell(t)}\sum_{n>t} Q(n) \le \limsup_{t\uparrow\infty} \ell(t)t^{\beta-1}e^{t^\beta\ell(t)}\sum_{n>t} Q(n) \le \frac{1+\epsilon}{\beta-\epsilon}.$$

From here, we obtain (G.2) by sending ϵ to zero.

H Proof of Theorem 3.2

The renewal equation defines a unique function p according to the scheme $p(1) := c_1/c_0$ and $p(t + 1) := c_{t+1}/c_0 - \sum_{s=1}^{t} p(s)c_{t-s+1}/c_0$ for $t \geq 1$. We must verify that $p(s) \geq 0$ for all $s \geq 1$ and that $\sum_{s \geq 1} p(s) = 1$.

We verify non-negativity of p by induction. We have $p(1) := c_1/c_0 > 0$. Pick $t \geq 1$ and assume that $p(s) \geq 0$ for every $s \leq t$. To say that c is a Kaluza sequence means to say that $\{c_n/c_{n-1}\}_{n \geq 1}$ is non-decreasing, so that $c_{t-s+1}/c_{t-s} \leq c_{t+1}/c_t$ for each $s \leq t$. Then, the renewal equation yields

$$p(t + 1) = c_{t+1}/c_0 - \sum_{s=1}^{t} p(s)\, c_{t-s+1}/c_0 \geq c_{t+1}[c_t - \sum_{s=1}^{t} p(s)\, c_{t-s}]/(c_0 c_t) = 0.$$

Let us now show that p sums to 1. By combining non-negativity of p with the renewal equation we get $\sum_{s=1}^{n} p(s)\, c_{t-s} \leq c_t$ for all $t \geq n \geq 1$. This bound gives $\sum_{s \geq 1} p(s) \leq 1$ by sending t to infinity and by recalling that $\lim_{t \uparrow \infty} c_t = c_0^2$ by hypothesis. Then, the fact that $\sum_{s \geq 1} p(s) < \infty$ allows us an application of the dominated convergence theorem to the renewal equation, which shows that $\sum_{s \geq 1} p(s) = 1$ by using $\lim_{t \uparrow \infty} c_t = c_0^2$ once again.

I Proof of Lemma 4.1

Since $\mathscr{B}$ is the σ-algebra generated by cylinder subsets of $\mathbb{N}_2^\infty$, Theorem 1.17 of [35] tells us that it suffices to verify the lemma for cylinder subsets. Pick two cylinder subsets $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{N}_2^\infty$, so that there exist integers $m \geq 1$ and $n \geq 1$ and sets $A \subseteq \mathbb{N}_2^m$ and $B \subseteq \mathbb{N}_2^n$ such that $\mathcal{A} = \{x \in \mathbb{N}_2^\infty : (x_1, \ldots, x_m) \in A\}$ and $\mathcal{B} = \{x \in \mathbb{N}_2^\infty : (x_1, \ldots, x_n) \in B\}$. By introducing the functions f on $\mathbb{N}_2^m$ and g on $\mathbb{N}_2^n$ defined respectively by $f(x_1, \ldots, x_m) := \mathbb{1}_{\{(x_1,\ldots,x_m) \in A\}} - \mathbb{1}_{\{(0,\ldots,0) \in A\}}$ and $g(x_1, \ldots, x_n) := \mathbb{1}_{\{(x_1,\ldots,x_n) \in B\}} - \mathbb{1}_{\{(0,\ldots,0) \in B\}}$, we can write for every $t \geq 1$

$$\mathbb{P}_o\big[\mathcal{A} \cap \mathcal{T}^{-t-m+1}\mathcal{B}\big] - \mathbb{P}_o\big[\mathcal{A}\big]\mathbb{P}_o\big[\mathcal{B}\big]$$
$$= \mathbb{P}\big[X \in \mathcal{A}, \mathcal{T}^{t+m-1}X \in \mathcal{B}\big] - \mathbb{P}\big[X \in \mathcal{A}\big]\mathbb{P}\big[X \in \mathcal{B}\big]$$
$$= \mathbb{P}\big[(X_1, \ldots, X_m) \in A, (X_{t+m}, \ldots, X_{t+m+n-1}) \in B\big]$$
$$\quad - \mathbb{P}\big[(X_1, \ldots, X_m) \in A\big]\mathbb{P}\big[(X_1, \ldots, X_n) \in B\big]$$
$$= \mathrm{cov}\big[f(X_1, \ldots, X_m), g(X_{m+t}, \ldots, X_{m+t+n-1})\big].$$

Since $f(0, \ldots, 0) = g(0, \ldots, 0) = 0$, we can invoke Lemma 3.1 to conclude that for all $t \geq 1$

$$\mathbb{P}_o\big[\mathcal{A} \cap \mathcal{T}^{-t-m+1}\mathcal{B}\big] - \mathbb{P}_o\big[\mathcal{A}\big]\mathbb{P}_o\big[\mathcal{B}\big]$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n} C_{i,j}(t)\, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=i\}} f(X_m, \ldots, X_1)]}{\mathbb{P}[S_1 = i]}\, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=j\}} g(X_1, \ldots, X_n)]}{\mathbb{P}[S_1 = j]}, \qquad \text{(I.1)}$$

where for $i \geq 1$ and $j \geq 1$

$$C_{i,j}(t) := \sum_{u=1}^{i}\sum_{v=1}^{j} p(i - u)\, \rho_{u+t+v-2}\, p(j - v)$$

with $p(0) := -1$. The only dependence on t is contained in the autocovariance. Formula I.1 proves the lemma since aperiodicity of p entails $\lim_{t\uparrow\infty} \rho_t = 0$.

J Proof of Corollary 4.1

Set $\mathcal{Y}_t := \{(x_1, \ldots, x_t) \in \mathbb{N}_2^t : |(\mu/t) \ln \pi_t(x_1, \ldots, x_t) + \mathcal{H}(p)| \le \epsilon\}$ for every integer $t \ge 1$. Theorem 4.1 yields $\lim_{t\uparrow\infty} \mathbb{P}[(X_1, \ldots, X_t) \in \mathcal{Y}_t] = 1$, so that $\mathbb{P}[(X_1, \ldots, X_t) \in \mathcal{Y}_t] \ge 1 - \delta$ for all sufficiently large t. On the other hand, the bounds $\sum_{(x_1, \ldots, x_t) \in \mathcal{Y}_t} \pi_t(x_1, \ldots, x_t) \le 1$ and $\pi_t(x_1, \ldots, x_t) \ge e^{-(t/\mu)[\mathcal{H}(p)+\epsilon]}$ for $(x_1, \ldots, x_t) \in \mathcal{Y}_t$ result in $|\mathcal{Y}_t| \le e^{(t/\mu)[\mathcal{H}(p)+\epsilon]}$. Thus, (i) is proved by taking $\mathcal{X}_o = \mathcal{Y}_t$.

Let us verify (ii). Pick a typical set $\mathcal{X}$ of length t and set $\mathcal{Y}_t := \{(x_1, \ldots, x_t) \in \mathbb{N}_2^t : |(\mu/t) \ln \pi_t(x_1, \ldots, x_t) + \mathcal{H}(p)| \le \epsilon/2\}$ for $t \ge 1$. Since $\mathbb{P}[(X_1, \ldots, X_t) \in \mathcal{X}] \ge 1 - \delta$ by definition with $\delta \in (0, 1)$ and $\lim_{t\uparrow\infty} \mathbb{P}[(X_1, \ldots, X_t) \in \mathcal{Y}_t] = 1$ by Theorem 4.1, for all sufficiently large t we have

$$
\begin{aligned}
\mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{Y}_t \cap \mathcal{X}\big] &= \mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{Y}_t\big] + \mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{X}\big] \\
&\quad - \mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{Y}_t \cup \mathcal{X}\big] \\
&\ge \mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{Y}_t\big] - \delta \ge e^{-(t/\mu)(\epsilon/2)}.
\end{aligned}
$$

At the same time

$$
\begin{aligned}
\mathbb{P}\big[(X_1, \ldots, X_t) \in \mathcal{Y}_t \cap \mathcal{X}\big] &= \sum_{(x_1, \ldots, x_t) \in \mathcal{Y}_t \cap \mathcal{X}} \pi_t(x_1, \ldots, x_t) \\
&\le |\mathcal{X}| e^{-(t/\mu)[\mathcal{H}(p) - \epsilon/2]}.
\end{aligned}
$$

These two bounds prove (ii).

K Proof of Proposition 4.1

Fix $t \ge 1$ and assume that $\sum_{n \ge t} |\rho_{n+1} - \rho_n| < \infty$ and $\sum_{n \ge t} n |\rho_{n+1} - 2\rho_n + \rho_{n-1}| < \infty$, otherwise there is nothing to prove. We verify first that for all positive integers m and n and all sets $\mathcal{A} \in \mathscr{F}_1^m$ and $\mathcal{B} \in \mathscr{F}_{m+t}^{m+t+n-1}$

$$
\Big| \mathbb{P}[\mathcal{A} \cap \mathcal{B}] - \mathbb{P}[\mathcal{A}]\, \mathbb{P}[\mathcal{B}] \Big| \le \sum_{i \ge 1} \sum_{j \ge 1} |C_{i,j}(t)|. \tag{K.1}
$$

Coefficients $C_{i,j}(t)$ were introduced by Lemma 3.1. A monotone class argument [36] allows us to extend the bound (K.1) to all $\mathcal{B} \in \mathscr{F}_{m+t}^{\infty}$, so that $\alpha(t) \le \sum_{i \ge 1} \sum_{j \ge 1} |C_{i,j}(t)|$. Then we demonstrate that

$$
\sum_{i \ge 1} \sum_{j \ge 1} |C_{i,j}(t)| \le 3\mu^2 \sum_{n \ge t} |\rho_{n+1} - \rho_n| + 4\mu^2 \sum_{n \ge t} n |\rho_{n+1} - 2\rho_n + \rho_{n-1}|. \tag{K.2}
$$

Let us verify (K.1). Pick $m \ge 1$, $n \ge 1$, $\mathcal{A} \in \mathscr{F}_1^m$, and $\mathcal{B} \in \mathscr{F}_{m+t}^{m+t+n-1}$. Since $\mathcal{A}$ is measurable with respect to $\mathscr{F}_1^m$, there exists a set $A \subseteq \mathbb{N}_2^m$ such that $\mathcal{A} = \{\omega \in \Omega : (X_1(\omega), \ldots, X_m(\omega)) \in A\}$. In the same way, there exists $B \subseteq \mathbb{N}_2^n$ with the property that $\mathcal{B} = \{\omega \in \Omega : (X_{m+t}(\omega), \ldots, X_{m+t+n-1}(\omega)) \in B\}$. Set $f(x_1, \ldots, x_m) := \mathbb{1}_{\{(x_1, \ldots, x_m) \in A\}} - \mathbb{1}_{\{(0, \ldots, 0) \in A\}}$ and $g(x_1, \ldots, x_n) := \mathbb{1}_{\{(x_1, \ldots, x_n) \in B\}} - \mathbb{1}_{\{(0, \ldots, 0) \in B\}}$. Then, $f(0, \ldots, 0) = g(0, \ldots, 0) = 0$ and Lemma 3.1 yields

$$\mathbb{P}[\mathcal{A} \cap \mathcal{B}] - \mathbb{P}[\mathcal{A}] \, \mathbb{P}[\mathcal{B}]$$
$$= \mathbb{P}\big[(X_1, \ldots, X_m) \in A, (X_{m+t}, \ldots, X_{m+t+n-1}) \in B\big]$$
$$\quad - \mathbb{P}\big[(X_1, \ldots, X_m) \in A\big] \mathbb{P}\big[(X_{m+t}, \ldots, X_{m+t+n-1}) \in B\big]$$
$$= \mathrm{cov}\big[f(X_1, \ldots, X_m), g(X_{m+t}, \ldots, X_{m+t+n-1})\big]$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} C_{i,j}(t) \, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=i\}} f(X_m, \ldots, X_1)]}{\mathbb{P}[S_1 = i]} \, \frac{\mathbb{E}[\mathbb{1}_{\{S_1=j\}} g(X_1, \ldots, X_n)]}{\mathbb{P}[S_1 = j]},$$

where for $i \geq 1$ and $j \geq 1$

$$C_{i,j}(t) := \sum_{u=1}^{i} \sum_{v=1}^{j} p(i-u) \, \rho_{u+t+v-2} \, p(j-v)$$

with $p(0) := -1$. As f and g can only take the values -1, 0, and 1, this formula results in

$$\left| \mathbb{P}[\mathcal{A} \cap \mathcal{B}] - \mathbb{P}[\mathcal{A}] \, \mathbb{P}[\mathcal{B}] \right| \leq \sum_{i=1}^{m} \sum_{j=1}^{n} |C_{i,j}(t)| \leq \sum_{i \geq 1} \sum_{j \geq 1} |C_{i,j}(t)|.$$

Let us move now to (K.2). Simple algebra shows that for all $i \geq 1$ and $j \geq 1$

$$C_{i,j}(t) = \rho_{i+t+j-2} - \sum_{u=1}^{i-1} p(i-u) \, \rho_{u+t+j-2} - \sum_{v=1}^{j-1} \rho_{i+t+v-2} \, p(j-v)$$
$$+ \sum_{u=1}^{i-1} \sum_{v=1}^{j-1} p(i-u) \, \rho_{u+t+v-2} \, p(j-v)$$
$$= \rho_{i+t+j-2} \, Q(i-1) Q(j-1)$$
$$+ \sum_{u=1}^{i-1} p(u) \left(\rho_{i+t+j-2} - \rho_{i+t+j-2-u} \right) Q(j-1)$$
$$+ \sum_{v=1}^{j-1} Q(i-1) \left(\rho_{i+t+j-2} - \rho_{i+t+j-2-v} \right) p(v)$$
$$+ \sum_{u=1}^{i-1} \sum_{v=1}^{j-1} p(u) \left(\rho_{i+t+j-2} - \rho_{i+t+j-2-v} - \rho_{i+t+j-2-u} + \rho_{i+t+j-2-u-v} \right) p(v).$$

We recall that Q is the tail of the probability distribution p. This identity leads to the bound

$$\sum_{i \geq 1} \sum_{j \geq 1} |C_{i,j}(t)| \leq K_1 + 2K_2 + K_3 \tag{K.3}$$

with

$$K_1 := \sum_{i \geq 0} \sum_{j \geq 0} |\rho_{i+t+j}| \, Q(i) Q(j),$$

$$K_2 := \sum_{i \geq 1} \sum_{j \geq 0} \sum_{u=1}^{i} p(u) \left| \rho_{i+t+j} - \rho_{i+t+j-u} \right| Q(j),$$

and

$$K_3 := \sum_{i \geq 1} \sum_{j \geq 1} \sum_{u=1}^{i} \sum_{v=1}^{j} p(u) \left| \rho_{i+t+j} - \rho_{i+t+j-v} - \rho_{i+t+j-u} + \rho_{i+t+j-u-v} \right| p(v).$$

We address K_1, K_2, and K_3 separately.

To begin with, we stress that the condition $\sum_{n \geq t} |\rho_{n+1} - \rho_n| < \infty$ implies $\lim_{n \uparrow \infty} \rho_n = 0$. In fact, if p is not aperiodic, then there exists an integer $\tau > 1$ such that $p(s) = 0$ unless s is a multiple of τ. The renewal equation gives $c_n = 0$ unless n is a multiple of τ, and the renewal theorem [31] shows that $\lim_{k \uparrow \infty} c_{k\tau} = 1/\mu$. In this situation, $|\rho_{k\tau+1} - \rho_{k\tau}| = |c_{k\tau+1} - c_{k\tau}| = |c_{k\tau}| \geq 1/(2\mu)$ for all sufficiently large k, which contradicts the hypothesis $\sum_{n \geq t} |\rho_{n+1} - \rho_n| < \infty$. Thus, $\sum_{n \geq t} |\rho_{n+1} - \rho_n| < \infty$ requires that p is aperiodic, and hence that $\lim_{n \uparrow \infty} \rho_n = 0$. The limit $\lim_{n \uparrow \infty} \rho_n = 0$ justifies the identity $\rho_{i+t+j} = \sum_{n \geq i+t+j} (\rho_n - \rho_{n+1})$, which yields the inequality $|\rho_{i+t+j}| \leq |\sum_{n \geq t} |\rho_{n+1} - \rho_n|$, which results in the bound

$$K_1 \leq \mu^2 \sum_{n \geq t} |\rho_{n+1} - \rho_n|. \tag{K.4}$$

As far as K_2 is concerned, by rearranging indices we get

$$K_2 = \sum_{u \geq 1} \sum_{l \geq t} p(u) |\rho_{l+u} - \rho_l| \sum_{j=0}^{l-t} Q(j) \leq \mu \sum_{u \geq 1} \sum_{l \geq t} p(u) |\rho_{l+u} - \rho_l|.$$

At this point, the bound $|\rho_{l+u} - \rho_l| \leq \sum_{n=l}^{l+u-1} |\rho_{n+1} - \rho_n|$ shows that

$$
\begin{aligned}
K_2 &\leq \mu \sum_{u \geq 1} \sum_{l \geq t} p(u) \sum_{n=l}^{l+u-1} |\rho_{n+1} - \rho_n| \\
&= \mu \sum_{n \geq t} \sum_{u \geq 1} (n - t + 1 \wedge u) \, p(u) \, |\rho_{n+1} - \rho_n| \\
&\leq \mu \sum_{n \geq t} \sum_{u \geq 1} u p(u) \, |\rho_{n+1} - \rho_n| = \mu^2 \sum_{n \geq t} |\rho_{n+1} - \rho_n|. \tag{K.5}
\end{aligned}
$$

Finally, let us consider K_3. By rearranging indices we can write

$$
\begin{aligned}
K_3 &= \sum_{l \geq t} \sum_{u \geq 1} \sum_{v \geq 1} (l - t + 1) \, p(u) \left| \rho_{l+u+v} - \rho_{l+u} - \rho_{l+v} + \rho_l \right| p(v) \\
&\leq 2 \sum_{l \geq t} l \sum_{u \geq 1} \sum_{v \geq u} p(u) \left| \rho_{l+u+v} - \rho_{l+u} - \rho_{l+v} + \rho_l \right| p(v).
\end{aligned}
$$

In order to go one step further, setting $\Delta_n := \rho_{n+1} - 2\rho_n + \rho_{n-1}$ for brevity, we resort to the identity

$$\rho_{l+u+v} - \rho_{l+u} - \rho_{l+v} + \rho_l = \sum_{k=v+1}^{u+v} (u + v - k) \, \Delta_{k+l} + \sum_{k=1}^{v} u \wedge k \, \Delta_{k+l},$$

which holds for $u \leq v$ as one can easily verify. This identity gives

$$K_3 \leq 2M_1 + 2M_2 \tag{K.6}$$

with

$$M_1 := \sum_{l \geq t} l \sum_{u \geq 1} \sum_{v \geq u} p(u) \sum_{k=v+1}^{u+v} (u + v - k) |\Delta_{k+l}| \, p(v)$$

and

$$M_2 := \sum_{l \geq t} l \sum_{u \geq 1} \sum_{v \geq u} p(u) \sum_{k=1}^{v} u \wedge k \, |\Delta_{k+l}| \, p(v).$$

By rearranging indices we can write

$$M_1 = \sum_{n \geq t+2} |\Delta_n| \sum_{u=1}^{n-t-1} \sum_{v=u}^{n-t-1} p(u) \, p(v) \sum_{l=t \vee (n-u-v)}^{n-v-1} l(u + v - n + l)$$

$$\leq \sum_{n \geq t+2} |\Delta_n| \sum_{u=1}^{n-t-1} \sum_{v=u}^{n-t-1} p(u) \, p(v) \, nu \big[n - v - t \vee (n - u - v) \big].$$

Under the condition $u \leq v$, if $n - u - v \geq t$, then $n - v - t \vee (n - u - v) = u \leq v$ and if $n - u - v < t$, then $n - v - t \vee (n - u - v) = n - v - t < u \leq v$. We thus have

$$M_1 \leq \sum_{n \geq t+2} |\Delta_n| \sum_{u=1}^{n-t-1} \sum_{v=u}^{n-t-1} p(u) \, p(v) \, nuv \leq \mu^2 \sum_{n \geq t} n \, |\Delta_n|. \qquad \text{(K.7)}$$

Moving to M_2, by rearranging indices we find

$$M_2 \leq \sum_{l \geq t} l \sum_{u \geq 1} \sum_{v \geq u} u p(u) \sum_{k=1}^{v} |\Delta_{k+l}| \, p(v)$$

$$= \sum_{n \geq t+1} |\Delta_n| \sum_{u \geq 1} \sum_{v \geq u} u p(u) \, p(v) \sum_{l=t \vee (n-v)}^{n-1} l$$

$$\leq \sum_{n \geq t+1} |\Delta_n| \sum_{u \geq 1} \sum_{v \geq u} u p(u) \, p(v) \, n \big[n - t \vee (n - v) \big].$$

If $n - v \geq t$, then $n - t \vee (n - v) = v$ and if $n - v < t$, then $n - t \vee (n - v) = n - t < v$. It follows that

$$M_2 \leq \sum_{n \geq t+1} |\Delta_n| \sum_{u \geq 1} \sum_{v \geq u} u p(u) \, p(v) \, nv \leq \mu^2 \sum_{n \geq t} n \, |\Delta_n|. \qquad \text{(K.8)}$$

Bound (K.2) follows by combining (K.3) with (K.4), (K.5), (K.6), (K.7), and (K.8).

L Proof of Lemma 4.2

Assume that p is aperiodic. To begin with, we demonstrate that there exists a real sequence $\{\gamma_t\}_{t \geq 0}$ such that $\sum_{t \geq 0} |\gamma_t| < \infty$ and

$$\frac{1}{\sum_{t \geq 0} Q(t) z^t} = \sum_{t \geq 0} \gamma_t z^t \qquad \text{(L.1)}$$

for all complex numbers z in the open unit disk, Q being the tail of the waiting time distribution p. Consider the function f defined by $f(z) := \sum_{t \geq 0} Q(t)z^t$ for $|z| \leq 1$. We have $1 - \sum_{s \geq 1} p(s)z^s = (1 - z)f(z)$ for all $|z| \leq 1$. We claim that f has no zeros for $|z| \leq 1$. In fact, $|z| < 1$ entails $|\sum_{s \geq 1} p(s)z^s| < 1$, so that any zeros of f must occur on the circle $|z| = 1$. The point $z = 1$ cannot be a zero since $f(1) = \mu \neq 0$. If $f(e^{i\theta}) = 0$ for some $\theta \in (0, 2\pi)$, then $\sum_{s \geq 1} p(s)e^{is\theta} = 1$, which gives $\cos(s\theta) = 1$ for all $s \geq 1$ such that $p(s) > 0$. But this is impossible because p is aperiodic. Then, the function that maps z in $1/f(z)$ has no singularities in the open unit disk and is continuous in the closed unit disk. As a consequence, for $|z| < 1$ it can be expanded in a power series as $1/f(z) = \sum_{t \geq 0} \gamma_t z^t$ with γ_t defined for every $t \geq 0$ by

$$\gamma_t := \frac{1}{2\pi i} \int_{|z|=r} \frac{z^{-t-1}}{f(z)} dz = \frac{1}{2\pi r^t} \int_0^{2\pi} \frac{e^{-it\theta}}{f(re^{i\theta})} d\theta.$$

Here r is any number in $(0, 1)$. As $1/f$ is bounded on the closed unit disk by continuity, we can invoke the dominated convergence theorem to set $r = 1$ in the last integral:

$$\gamma_t = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-it\theta}}{f(e^{i\theta})} d\theta. \tag{L.2}$$

At the same time, since $f(e^{i\theta}) = \sum_{t \geq 0} Q(t)e^{it\theta}$ converges absolutely and has no zeros for real values of θ, Theorem 18.21 of [42] states that $1/f(e^{i\theta}) = \sum_{t=-\infty}^{\infty} \xi_t e^{it\theta}$ for all $\theta \in [0, 2\pi]$ with coefficients that fulfills $\sum_{t=-\infty}^{\infty} |\xi_t| < \infty$. By plugging this expansion in (L.2) we get $\gamma_t = \xi_t$ for every $t \geq 0$, and hence $\sum_{t \geq 0} |\gamma_t| < \infty$.

We now observe that by taking a Cauchy product, formula (L.1) shows that the sequence $\{\gamma_t\}_{t \geq 0}$ solves the problem $\gamma_0 = 1$ and $\gamma_t = -\sum_{s=1}^{t} Q(s)\gamma_{t-s}$ for all $t \geq 1$. Simple manipulations of the renewal equation demonstrate that the same problem is solved by the sequence $\{\gamma'_t\}_{t \geq 0}$ with entries $\gamma'_0 = 1$ and $\gamma'_t = \mu(\rho_t - \rho_{t-1})$ for $t \geq 1$. Since this problem has a unique solution, we get $\gamma_t = \mu(\rho_t - \rho_{t-1})$ for $t \geq 1$ and

$$\frac{1}{\sum_{t \geq 0} Q(t)z^t} = 1 + \mu \sum_{t \geq 1} (\rho_t - \rho_{t-1})z^t$$

for every $|z| < 1$. Derivative with respect to z gives

$$\frac{\sum_{t \geq 1} t Q(t)z^t}{[\sum_{t \geq 0} Q(t)z^t]^2} = \mu \sum_{t \geq 1} t(\rho_{t-1} - \rho_t)z^t. \tag{L.3}$$

This formula is what we need to prove the equivalence between points (i) and (ii) of the lemma. In fact, if p is aperiodic and $\sum_{s \geq 1} s^2 p(s) < \infty$, then $1/\sum_{t \geq 0} Q(t)z^t = \sum_{t \geq 0} \gamma_t z^t$ with $\sum_{t \geq 0} |\gamma_t| < \infty$ and $\sum_{t \geq 1} t Q(t) = \sum_{s \geq 1} (1/2)(s-1)sp(s) < \infty$. This way, since the Cauchy product of absolutely convergent series is absolutely convergent, by expanding the l.h.s. of (L.3) in a power series and by making a comparison with the r.h.s. we realize that $\sum_{t \geq 1} t|\rho_t - \rho_{t-1}| < \infty$. If instead $\sum_{t \geq 1} t|\rho_t - \rho_{t-1}| < \infty$, then p is aperiodic as we have seen in the proof of Proposition 4.1, so that (L.3) holds. It follows that $\sum_{t=1}^{n} t Q(t) \leq \mu^3 \sum_{t \geq 1} t|\rho_t - \rho_{t-1}| < \infty$ for all positive n, which suffices to prove that $\sum_{s \geq 1} s^2 p(s) < \infty$.

Suppose now that p is aperiodic and that $\sum_{s \geq 1} s^2 p(s) < \infty$. Let us demonstrate that $\sum_{t \geq 1} \alpha_t < \infty$. Proposition 4.1 tells us that $\sum_{t \geq 1} \alpha_t \leq 3\mu^2 \sum_{t \geq 1} t|\rho_t - \rho_{t-1}| + 4\mu^2 \sum_{t \geq 1} t^2|\rho_{t+1} - 2\rho_t + \rho_{t-1}|$. We already know that $\sum_{t \geq 1} t|\rho_t - \rho_{t-1}| < \infty$ and it remains to verify that $\sum_{t \geq 1} t^2|\rho_{t+1} - 2\rho_t + \rho_{t-1}| < \infty$. By taking the derivative of (L.3) with respect to z we get

$$\frac{\sum_{t\geq 1} t^2 Q(t) z^t}{[\sum_{t\geq 0} Q(t) z^t]^2} - 2\frac{[\sum_{t\geq 1} t Q(t) z^t]^2}{[\sum_{t\geq 0} Q(t) z^t]^3} = \mu \sum_{t\geq 1} t^2 (\rho_{t-1} - \rho_t) z^t. \qquad (L.4)$$

A shift of the index of the last series yields

$$\frac{\sum_{t\geq 1} t^2 Q(t) z^{t-1}}{[\sum_{t\geq 0} Q(t) z^t]^2} - 2z\frac{[\sum_{t\geq 1} t Q(t) z^{t-1}]^2}{[\sum_{t\geq 0} Q(t) z^t]^3} = \mu \sum_{t\geq 0} (t+1)^2 (\rho_t - \rho_{t+1}) z^t. \qquad (L.5)$$

By subtracting (L.5) from (L.4) we reach the result

$$\frac{\sum_{t\geq 1} t^2 p(t+1) z^t + \sum_{t\geq 1} (1-2t) Q(t) z^{t-1}}{[\sum_{t\geq 0} Q(t) z^t]^2} + 2z(1-z)\frac{[\sum_{t\geq 1} t Q(t) z^{t-1}]^2}{[\sum_{t\geq 0} Q(t) z^t]^3}$$

$$= \mu \sum_{t\geq 1} t^2 (\rho_{t-1} - 2\rho_t + \rho_{t+1}) z^t + \mu \sum_{t\geq 1} (1-2t)(\rho_{t-1} - \rho_t) z^{t-1}. \qquad (L.6)$$

As before, using $1/\sum_{t\geq 0} Q(t) z^t = \sum_{t\geq 0} \gamma_t z^t$ and bearing in mind that the Cauchy product of absolutely convergent series is absolutely convergent, by expanding the l.h.s. of (L.6) in a power series and by making a comparison with the r.h.s. we find $\sum_{t\geq 1} t^2 |\rho_{t-1} - 2\rho_t + \rho_{t+1}| < \infty$.

At last, let us verify the formula $\sum_{s\geq 1} s^2 p(s) = \mu + 2\mu^3 \sum_{t\geq 0} \rho_t$. If p is aperiodic and $\sum_{s\geq 1} s^2 p(s) < \infty$, or equivalently if $\sum_{t\geq 1} t|\rho_t - \rho_{t-1}| < \infty$, then we can apply Abel's theorem to (L.3) to obtain $\sum_{s\geq 1} s^2 p(s) = \mu + 2\mu^3 \sum_{t\geq 1} t(\rho_{t-1} - \rho_t)$. We have $\sum_{t\geq 0} |\rho_t| < \infty$ since $\rho_t = \sum_{n>t} (\rho_{n-1} - \rho_n)$ as $\lim_{n\uparrow\infty} \rho_n = 0$ and $\sum_{t\geq 1} t|\rho_t - \rho_{t-1}| < \infty$. Thus, the identity $\sum_{t=1}^{n} t(\rho_{t-1} - \rho_t) = \sum_{t=0}^{n-1} \rho_t - n\rho_n$ for $n \geq 1$ tells us that $\lim_{n\uparrow\infty} n|\rho_n|$ exists, and the fact that $\sum_{t\geq 0} |\rho_t| < \infty$ implies $\lim_{n\uparrow\infty} n|\rho_n| = 0$. At this point, the same identity $\sum_{t=1}^{n} t(\rho_{t-1} - \rho_t) = \sum_{t=0}^{n-1} \rho_t - n\rho_n$ for $n \geq 1$ leads to $\sum_{t\geq 1} t(\rho_{t-1} - \rho_t) = \sum_{t\geq 0} \rho_t$, which proves that $\sum_{s\geq 1} s^2 p(s) = \mu + 2\mu^3 \sum_{t\geq 0} \rho_t$.

References

1. Nguyen, H.C., Zecchina, R., Berg, J.: Inverse statistical problems: from the inverse Ising problem to data science. Adv. Phys. **66**, 197–261 (2017)
2. De Jong, H.: Modeling and simulation of genetic regulatory systems: a literature review. J. Comput. Biol. **9**, 67–103 (2002)
3. ...Kass, R.E., Amari, S.I., Arai, K., Brown, E.N., Diekman, C.O., Diesmann, M., Doiron, B., Eden, U.T., Fairhall, A.L., Fiddyment, G.M., Fukai, T., Grün, S., Harrison, M.T., Helias, M., Nakahara, H., Teramae, J., Thomas, P.J., Reimers, M., Rodu, J., Rotstein, H.G., Shea-Brown, E., Shimazaki, H., Shinomoto, S., Yu, B.M., Kramer, M.A.: Computational neuroscience: mathematical and statistical perspectives. Annu. Rev. Stat. Appl. **5**, 183–214 (2018)
4. Frees, E.W.: Regression modeling with actuarial and financial applications. Cambridge University Press, New York (2010)
5. Qin, H., Gong, R., Liu, X., Bai, X., Song, J., Sebe, N.: Binary neural networks: a survey. Pattern Recognit. **105**, 107281 (2020)
6. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. **1**, 1–305 (2008)
7. Fokianos, F.K., Kedem, B.: Regression models for time series analysis. Wiley, New Jersey (2002)
8. Usatenko, O.V., Yampol'skii, V.A.: Binary N-step Markov chains and long-range correlated systems. Phys. Rev. Lett. **90**, 110601 (2003)
9. Usatenko, O.V., Yampol'skii, V.A., Kechedzhy, K.E., Melnyk, S.S.: Symbolic stochastic dynamical systems viewed as binary N-step Markov chains. Phys. Rev. E **68**, 061107 (2003)
10. Hod, S., Keshet, U.: Phase transition in random walks with long-range correlations. Phys. Rev. E **70**, 015104(R) (2004)

11. Melnyk, S.S., Usatenko, O.V., Yampol'skii, V.A., Golick, V.A.: Competition between two kinds of correlations in literary texts. Phys. Rev. E **72**, 026140 (2005)
12. Melnyk, S.S., Usatenko, O.V., Yampol'skii, V.A.: Memory functions of the additive Markov chains: applications to complex dynamic systems. Physica A **361**, 405–415 (2006)
13. Melnyk, S.S., Usatenko, O.V., Yampol'skii, V.A., Apostolov, S.S., Maiselis, Z.A.: Memory functions and correlations in additive binary Markov chains. J. Phys. A: Math. Gen. **39**, 14289–14301 (2006)
14. Izrailev, F.M., Krokhin, A.A., Makarov, N.M., Melnyk, S.S., Usatenko, O.V., Yampol'skii, V.A.: Memory function versus binary correlator in additive Markov chains. Physica A **372**, 279–297 (2006)
15. Lomnicki, Z.A., Zaremba, S.K.: Applications of zero-one processes. J. R. Stat. Soc. Series B Stat. Methodol. **17**, 243–255 (1955)
16. Keenan, D.M.: A time series analysis of binary data. J. Am. Stat. Assoc. **77**, 816–821 (1982)
17. Izrailev, F.M., Krokhin, A.A., Makarov, N.M., Usatenko, O.V.: Generation of correlated binary sequences from white noise. Phys. Rev. E **76**, 027701 (2007)
18. Apostolov, S.S., Izrailev, F.M., Makarov, N.M., Mayzelis, Z.A., Melnyk, S.S., Usatenko, O.V.: The signum function method for the generation of correlated dichotomic chains. J. Phys. A: Math. Theor. **41**, 175101 (2008)
19. Usatenko, O.V., Melnik, S.S., Apostolov, S.S., Makarov, N.M., Krokhin, A.A.: Iterative method for generating correlated binary sequences. Phys. Rev. E **90**, 053305 (2014)
20. Sung, C.L., Hung, Y., Rittase, W., Zhu, C., Wu, J.: A generalized Gaussian process model for computer experiments with binary time series. J. Am. Stat. Assoc. **115**, 1–24 (2019)
21. Bradley, R.C.: Basic properties of strong mixing conditions. A survey and some open questions. Probab. Surv. **2**, 107–144 (2005)
22. Emrich, L.J., Piedmonte, M.R.: A Method for generating high-dimensional multivariate binary variates. Am. Stat. **45**, 302–304 (1991)
23. Serinaldi, F., Lombardo, F.: General simulation algorithm for autocorrelated binary processes. Phys. Rev. E **95**, 023312 (2017)
24. Serinaldi, F., Lombardo, F.: BetaBit: a fast generator of autocorrelated binary processes for geophysical research. Europhys. Lett. **118**, 30007 (2017)
25. Jiang, W., Song, S., Hou, L., Zhao, H.: A set of efficient methods to generate high-dimensional binary data with specified correlation structures. Am. Stat. 1–13 (2020)
26. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer, New York (2003)
27. Kingman, J.F.C.: Regenerative Phenomena. Wiley, London (1972)
28. Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957)
29. Jaynes, E.: Information theory and statistical mechanics. II. Phys. Rev. **108**, 171–190 (1957)
30. Galves, A., Löcherbach, E.: Stochastic chains with memory of variable length. TICSP Series **38**, 117–133 (2008)
31. Erdös, P., Feller, W., Pollard, H.: A property of power series with positive coefficients. Bull. Am. Math. Soc. **55**, 201–204 (1949)
32. Chover, J., Ney, P., Wainger, S.: Functions of probability measures. J. Anal. Math. **26**, 255–302 (1972)
33. Embrechts, P., Omey, E.: Functions of power series. Yokohama Math. J. **32**, 77–88 (1984)
34. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press, Cambridge (1989)
35. Walters, P.: An Introduction to Ergodic Theory. Springer, New York (1982)
36. Shiryaev, A.N.: Probability, 2nd edn. Springer, New York (1996)
37. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, New Jersey (2006)
38. Zamparo, M.: Large deviations in renewal models of statistical mechanics. J. Phys. A: Math. Theor. **52**, 495004 (2019)
39. Zamparo, M.: Large deviations in discrete-time renewal theory. Stoch. Process. Their Appl. **139**, 80–109 (2021)
40. Ibragimov, I.A., Linnik, Y.V.: Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen (1971)
41. Berbee, H.: Convergence rates in the strong law for bounded mixing sequences. Probab. Theory Relat. Fields **74**, 255–270 (1987)
42. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)