Long range dependent Markov chains with applications

Barlas Oğuz, Venkat Anantharam

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Email: {barlas, ananth}@eecs.berkeley.edu

***Abstract*—We discuss functions of long range dependent Markov chains. We state sufficient conditions under which an instantaneous function of a long range dependent Markov chain has the same Hurst index as the underlying chain. We discuss several applications of the theorem in the fields of information theory, queuing networks, and finance.**

I. Introduction

Although traditional stochastic models such as i.i.d., Poisson, or finite state Markov models are preferred for their simplicity and tractability, the fact remains that real systems often exhibit more complex and persistent memory than what can be captured by these models. Sometimes the memory effects are so persistent that the effects cannot be ignored even in asymptotic discussions. In these cases one needs to use long range dependent models.

A stationary random process (X_n) with $E[X_n^2] < \infty$ is said to be long range dependent (LRD) if

$$\limsup_{n\to\infty} \sum_{r=1}^{n} \mathrm{cov}(X_0, X_r) = \infty. \tag{1}$$

The degree of long range dependence is measured by the Hurst index H ($\frac{1}{2} \le H \le 1$).

$$H := inf \left\{ h : \limsup_{n\to\infty} \frac{\sum_{r=1}^{n} \mathrm{cov}(X_0, X_r)}{n^{2h-1}} < \infty \right\}.$$

Equivalently, we can write

$$H := inf \left\{ h : \limsup_{n\to\infty} \frac{\mathrm{var}(\sum_{i=1}^{n} X_i)}{n^{2h}} < \infty \right\}.$$

A process that is not LRD is said to be short range dependent (SRD). The justification for this division is as follows. Although SRD processes may have memory, the effect of this memory can be ignored in asymptotic discussions by taking long blocks of the original source and treating these blocks as a meta process. If the process is SRD, for many practical purposes, the block process can be well approximated by an i.i.d. process. As a result, SRD processes behave similarly to an i.i.d. process in many asymptotic settings. Indeed, the complement of the condition in 1 is necessary for a central limit theorem to hold.

On the other hand, the effect of memory in an LRD process does not disappear even asymptotically under any scaling. LRD processes do not satisfy the central limit theorem, and the limit of their scaled sums is not in general Gaussian. In fact, as the definitions suggest, the scaling required to obtain a meaningful limit is different than $\sqrt{n}$, and is related to the Hurst index of the process. The limiting processes, when they exist, are described by stable distributions and self similar processes. For a detailed description of these results and other properties of LRD processes, the reader is referred to the books [1],[2].

In the continuous time setting the most popular model that is used is the fractional Brownian motion (fBm) (see [1], chapter 7.2). In the discrete time case fractional ARIMA models have been widely adopted (see e.g. [3], chapter 2.5). Although parametric models such as these have their advantages in terms of model fitting and estimation, in many cases they can only provide an approximation to the underlying system. Here we will work with models based on countable state, long range dependent Markov chains, which is a much more flexible class of models. We introduce LRD Markov chains and some basic properties in section II. We will then explore applications through simple examples.

We will be looking at the use of LRD models in the fields of finance, queuing networks, and source coding. In these fields, there is sometimes enough empirical evidence for the presence of long range dependence that trading the traditional stochastic models for more complicated but realistic LRD ones has been warranted. The first use of an LRD model in finance is credited to Mandelbrot [4]. In section IV we will be examining an example based on his simple model for wheat prices. For an extensive survey of long range dependent models in finance, please see [5].

Interest in LRD processes in communication networks was sparked by several empirical observations that showed such distributions were characteristic of network traffic on the internet [6],[7],[8]. Due to the fundamentally different qualities of LRD processes mentioned above, these discoveries have important, and often negative consequences for the modeling and analysis of communication networks. Among these are different asymptotics for queue sizes and packet drop probabilities [9], [10], [11], [12], [13], [14], and a need for new optimal schedulers [15],[16],[17]. The mostly degrading effect of LRD traffic in networks has led to research efforts for understanding the mechanisms by which such traffic is generated and whether preventive measures are possible [16],[7]. To illustrate the use of LRD Markov models in networks, in section II-A we discuss a simple queuing network of two parallel queues, one of them being driven by an LRD process. We will show that under a fixed rate shared server with longest queue first scheduling, the busy-idle process of both queues will become LRD, (see also [18]).

In section III we will present an application to source coding. The relevance of LRD models to source coding stems from the work on variable bit-rate video. Variable bit-rate traffic (mainly VBR video) is an important component of internet traffic. In the hope of understanding such traffic better, there has been considerable work on analyzing traces of VBR video ([19], [20], [21], [22] to cite a few). The common observation that culminated from this work is that long range dependence is omnipresent in VBR traffic, and persists across a wide variety of codecs. Motivated by this observation, one wonders if long range dependence is an intrinsic property of certain information sources, independent of the coding algorithm chosen to encode the source. Example III contains a theorem that justifies this intuition in the context of coding a long range dependent renewal process.

The proofs in these examples all rely on a recently developed theorem about functions of LRD Markov chains [23][1]. The main result in this work, given in section V, provides us with sufficient conditions under which a function of an LRD Markov chain has the same Hurst index as the Markov chain itself. Our approach will involve first representing the problem as an LRD Markov chain, and then writing process of interest as a function of this chain. Then, using the theorem, we deduce the Hurst index of the relevant quantity from that of the underlying Markov chain. All proofs are deferred to section VI.

II. Long range dependent Markov models

Take (M_n), a positive-recurrent, aperiodic, discrete time, countable state Markov chain with state space $\mathbb{N}$, where $\mathbb{N}$ denotes the set of natural numbers. The chain is in stationarity with stationary distribution π. Long range dependence of a Markov chain on a general state space is defined in terms of the long range dependence of the indicator function $1(M_n = i)$ of state i of the chain. This function is LRD if and only if the indicator function of every state is LRD [24]. When this is true, (M_n) is said to be an LRD Markov chain. Moreover, the Hurst index of these functions is also a class property [24]. The common Hurst index H is said to be the Hurst index of the chain.

In [24] it is proved that a Markov chain is LRD if and only if the return time distribution of any state has infinite variance, i.e. that this happens if and only if the indicator function $1(M_n = i)$ is LRD. It is also argued that finite weighted sums of indicator functions on this chain also inherit this property. It is natural to conjecture that this might be true for all functions of the chain. However, this conjecture is disproved, most easily by considering a constant function (also see the two counterexamples in [24]). It is then of considerable interest to find which functions of an LRD Markov chain are also LRD. The main theorem in section V addresses this issue. We now turn to examples.

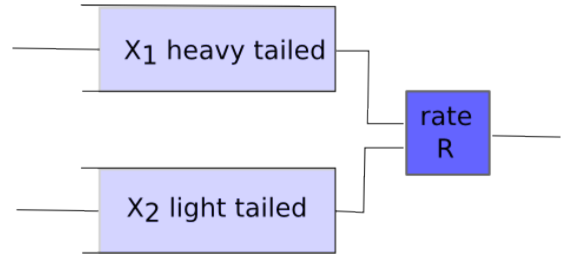


Fig. 1. Parallel queues with fixed a rate server.

A. Example 1: Longest queue first with mixed heavy and light tailed inputs

This example replicates the conclusion in [18] that long range dependence might spread under longest queue first (LQF) scheduling in a parallel queue setting.

There is a single server of rate $R \in \mathbb{N}$ with 2 parallel queues. The queues are fed by independent random processes, each modeled by a discrete time, irreducible, countable state Markov chain. As an example, we investigate the scenario where X_1 is i.i.d. with heavy tailed ($\mathrm{var}(X_1) = \infty$) arrival distribution on $\mathbb{N}$. $X_2 \in \mathbb{N}$ is either an i.i.d. process with light tailed ($\mathrm{var}(X_2) < \infty$) arrivals or a finite state $\mathbb{N}$-valued Markov chain in stationarity. We assume $E[X_1(0)] + E[X_2(0)] < R$.

Let $Q_1(n), Q_2(n)$ be the stationary queue lengths. We assume that the queue is work conserving, and moreover the scheduling decision at time n (number of packets to be served from each queue at time slot n) is a function of $(Q_1(n), Q_2(n))$, the queue sizes at time n. Given such a scheduling strategy, it is easily verified that $(X_1(n), X_2(n), Q_1(n), Q_2(n))$ is a countable state Markov chain.

Lemma II.1. *$(X_1(n), X_2(n), Q_1(n), Q_2(n))$ is positive recurrent.*

Proof: $E[X_1(0)] + E[X_2(0)] < R$ implies that the queue process $(Q_1(n), Q_2(n))$ is positive recurrent. Pick $M_1 > 0$ and define the set $S_1 = \{Q_1(n) + Q_2(n) < M_1\}$. The return times to this set have finite mean (say ν). Also define $S_2 = \{X_1(n) + X_2(n) < M_2\}$ (or in the case X_2 is a finite state chain, $S_2 = \{X_1(n) < M_2\}$) where M_2 is large enough such that S_2 is nonempty. $S_1 \cap S_2$ is a nonempty compact set. We claim the return times to this set have a finite mean. Since $1_n(S_2)$ is i.i.d, there is a positive probability (say at least p) of visiting S_2 each time there is a visit to S_1 (independent of previous visits). It is easily seen that the mean return time to $S_1 \cap S_2$ is at most ν/p. ■

We will look at long range dependence through the Hurst indices of the busy-idle processes of the queues. Let (X_1, Q_1')

[1]The technical content summarized in this paper was published in the Journal of Applied Probability vol.49 no.2 ©by Applied Probability Trust.

be the Markov chain if all the capacity were to be allocated to queue 1. Denote by $1(Q_1'(n) = 0)$, the busy-idle process of this queue. We know that the busy periods of Q_1' have infinite variance (see e.g. [25] theorem 8.10.3). Therefore both the Markov chain (X_1, Q_1') and the function $1(Q_1'(n) = 0)$ are LRD (see the beginning of this section). (X_2, Q_2'), similarly defined, is a short range dependent chain.

Lemma II.2. *$(X_1(n), X_2(n), Q_1(n), Q_2(n))$ is LRD.*

Proof: Consider the chain $(X_1(n), Q_1'(n), X_2(n), Q_2'(n))$. This chain is LRD because it is a combination of two independent chains (X_1, Q_1') and (X_2, Q_2'), one of which is LRD. Let t_1 be the return time to a nonempty compact set $S_1 = \{X_1(n), Q_1(n), X_2(n), Q_2(n) < M\}$. Similarly t_2 is the return time to the set $S_2 = \{X_1(n), Q_1'(n), X_2(n), Q_2'(n) < M\}$. Since $Q_1'(n) \leq Q_1(n)$ and $Q_2'(n) \leq Q_2(n)$, t_1 stochastically dominates t_2, and therefore $(X_1(n), X_2(n), Q_1(n), Q_2(n))$ is also LRD. ■

The question we want to ask then is whether $1(Q_2(n) = 0)$, the busy-idle process of the second queue (fed by short range dependent traffic), is also long range dependent. We claim:

Theorem II.3. *Under LQF scheduling, $1(Q_2(n) = 0)$ is LRD with the same Hurst index as $(X_1(n), X_2(n), Q_1(n), Q_2(n))$.*

It is clear that $1(Q_2(n) = 0)$ is a function of LRD Markov chain $(X_1(n), X_2(n), Q_1(n), Q_2(n))$. Therefore the problem fits within the framework of theorem V.

III. Example 2: Compressing a long range dependent renewal process

This section provides an alternative proof for the result in [26].

Let (X_n) be a discrete, stationary, ergodic renewal process. We begin by introducing the function

$$\rho_n(X_{-\infty}^n) := -\log P(X_n | X_{-\infty}^{n-1}),$$

which is of central importance to coding theory. The behavior of (ρ_n) restricts the minimum code length of lossless compression algorithms by the following lemma, [27], which is also proved in [28].

Lemma III.1 (Barron's Lemma)**.** *Given $\{c(n), n \geq 1\}$, positive constants with $\sum_n 2^{-c(n)} < \infty$, we have*

$$L_n(X_1^n) \geq -\log P(X_1^n | X_{-\infty}^0) - c(n), \text{ eventually, a.s. } . \quad (2)$$

Here $L_n(X_1^n)$ is the code length for the first n symbols of the source for some lossless coding algorithm that produces bit strings. $c(n)$ can be made logarithmic in n.

By the ergodic theorem, the limit of $\frac{1}{n}\sum_{i=1}^n \rho_i$ as $n \to \infty$ exists a.s. and equals $\eta := E[-\log P(X_1 | X_{-\infty}^0)]$, i.e. the entropy rate of (X_n). This implies the following well known first order converse source coding theorem for such sources.

Theorem III.2.

$$\liminf_n \frac{1}{n} L_n(X_1^n) \geq \eta, \text{ a.s. } .$$

Lemma III.1 is strong enough to permit second order refinements to theorem III.2 once we know more about the process (ρ_n). For example, in [28], it is shown that for certain short range dependent classes of sources (e.g. finite state Markov chains), and appropriate coding schemes (e.g. Lempel-Ziv coding), $(L_n - n\eta)$ satisfies a central limit theorem.

Here, we state a second order converse source coding theorem, stating that the bit length process (L_n) will eventually dominate a long range dependent process the growth of whose variance is identical to that of (X_n), so that, in particular, it has the same Hurst exponent as (X_n).

Theorem III.3. *Let (X_n) be an aperiodic, long range dependent, stationary, ergodic renewal process. Then, there exists a long range dependent random process (γ_n) such that*

$$L_n(X_1^n) \geq \gamma_n, \text{ eventually, a.s.}$$

for all uniquely decodable source codes. Moreover, (γ_n) has the same Hurst index as (X_n).

This immediately follows from Barron's lemma once we show (ρ_n) are LRD with the same Hurst index as (X_n). This will follow from theorem V.2 if we can set up (ρ_n) as a function of a Markov chain.

IV. Example 3: Long range dependence in financial time series

Let $(P_n, -\infty < n < \infty)$ be the price of some financial asset, and $X_n = \log P_n$. It is an established assumption that the log returns, $r_n = X_n - X_{n-1}$ is well modeled by a martingale difference process. Such a model accounts for the fact that the log returns exhibit little correlation. Nevertheless, it is also a widely observed fact that some instantaneous functions of the log returns, such as $|r_n|^d$, exhibit long memory, (see e.g. [29]).

The popular approach to modeling this behavior has been to explicitly write the dependence of the absolute log returns into the statistical description of the model. The result is the various long-memory autoregressive conditional heteroskedasticity (ARCH) process models of financial time series, ([30] for an example).

We want to show in this example that, given a martingale difference sequence (r_n) that can be represented as a function of a long range dependent Markov chain, the outcome that $|r_n|^d$ will exhibit long range dependence should not be considered surprising.

We want to illustrate this with a very simple example based on Mandelbrot's model for wheat prices ([4]). We should note that this simple model is for purposes of illustration only, and does not account for all known properties of financial time series. For instance, it has been observed in many situations that (r_n) has a finite variance, despite having a polynomially decaying marginal distribution. The (r_n) in this example has infinite variance. Nevertheless, the proof scheme used here to establish the long range dependence of $|r_n|^d$ should be applicable much more generally.

Let (W_n) be a stationary random process which models the weather. (W_n) can take on 3 values: good, bad, and neutral $\{g, b, n\}$. The length of a good period, T (number of consecutive good days), has the same distribution as the length of a bad or a neutral period. Let $P(T \geq t) = t^{-\alpha}$. T has finite mean but infinite variance (i.e. $1 < \alpha \leq 2$). A good or bad period is followed necessarily by a neutral period. A neutral period is followed by a good or bad period with equal probabilities.

Let $\hat{X}_n$ be the fundamental (log) price of the asset (which can be thought of as summarizing exogenous variables that affect the real price). $\hat{X}_n$ varies as follows: increases by 1 for every good day, decreases by 1 for every bad day, and stays the same for every neutral day. The market calculates the real (log) price by projecting the expected future fundamental price: $X_n = \lim_{t\to\infty} E[\hat{X}_{n+t}|\hat{X}^n_{-\infty}]$.

By construction, (r_n) itself is a martingale difference sequence. We will now show that:

Theorem IV.1. *$\rho_n = |r_n|^d$ is LRD with Hurst index $\frac{1}{2}(3-\alpha)$. ($0 < d < \alpha/2$ for var(ρ_0) to be finite).*

V. Main theorem

We will state the main result from [23]. First some notation: (M_n) is a positive-recurrent, discrete time, countable state Markov chain with state space $\mathbb{N}$ and stationary distribution π_i, $i \in \mathbb{N}$. Most of the notation we use is borrowed from [31].

$\rho : \mathbb{N} \to \mathbb{R}$ is such that $\sum_{i\in\mathbb{N}} \rho(i)^2 \pi_i < \infty$.

$\rho_n := \rho(M_n)$.

$\mu := \sum_i \rho(i)\pi_i$, is the mean of ρ.

$p^{(n)}_{ij} := P(M_n = j|M_0 = i), n \geq 0$.

${}_kp^{(n)}_{ij} := P(M_n = j; M_l \neq k, 0 < l < n|M_0 = i), n > 0$.

${}_kp^{*}_{ij} := \sum_{n=1}^{\infty} {}_kp^{(n)}_{ij}$.

${}_{\mathcal{H}}p^{(n)}_{ij} := P(M_n = j; M_l \notin \mathcal{H}, 0 < l < n|M_0 = i), n > 0$.

${}_{\mathcal{H}}p^{*}_{ij} := \sum_{n=1}^{\infty} {}_{\mathcal{H}}p^{(n)}_{ij}$.

$f^{(n)}_{ij} := {}_jp^{(n)}_{ij}, n > 0$.

$Q^{(n)}_{ij} := \sum_{r=1}^{n}(p^{(r)}_{ij} - \pi_j), n > 0$.

$R^{(n)}_{ij} := \sum_{r=1}^{n} Q^{(r)}_{ij}, n > 0$.

$T_j := \inf_t\{t > 0 : M_t = j\}$.

$m_{ij} := E_i[T_j]$.

$H := inf\left\{h : \limsup_{n\to\infty} \frac{\text{var}(\sum_{i=1}^{n} \mathbf{1}(M_i=1))}{n^{2h}} < \infty\right\}$, the Hurst index of (M_n).

$H_\rho := inf\left\{h : \limsup_{n\to\infty} \frac{\text{var}(\sum_{i=1}^{n} \rho_i)}{n^{2h}} < \infty\right\}$, the Hurst index of (ρ_n).

The following lemma is useful to know.

Lemma V.1. *For an LRD Markov chain,*

$$\lim_{n\to\infty} Q^{(n)}_{ij} = \infty, \tag{3}$$

$$\lim_{n\to\infty} \frac{R^{(n)}_{ij}}{n} = \infty, \tag{4}$$

$$\lim_{n\to\infty} \frac{Q^{(n)}_{ij}/\pi_j}{Q^{(n)}_{11}/\pi_1} = 1. \tag{5}$$

Proof: (5) is eq. 8 in [24]. (3) follows from eqs. 8 and 5 of [24]. (4) follows from (3). ■

The following theorem provides a condition under which (ρ_n) inherits the Hurst index of (M_n).

Theorem V.2. *Let $\{\mathcal{A}_k\}$, $1 \leq k \leq K$, be a finite partition of the state space $\mathbb{N}$.*

(condition 1) Let $\mathcal{H}$ be a non-empty finite set, and

$$\lim_{n\to\infty} \frac{1}{Q^{(n)}_{11}/\pi_1} \sum_{r=1}^{n} \sum_{i\in\mathcal{A}_k, j\in\mathcal{A}_l} \pi_i |\rho(i)-\mu||\rho(j)-\mu| {}_{\mathcal{H}}p^{(r)}_{ij} = 0, \quad \forall k \neq l.$$

Also suppose $\pi^{\infty}_{\mathcal{A}_k} := \lim_{n\to\infty} \frac{\sum_{i,j\in\mathcal{A}_k} \pi_i \sum_{r=1}^{n} {}_1p^{(r)}_{ij}}{\sum_{i,j} \pi_i \sum_{r=1}^{n} {}_1p^{(r)}_{ij}}$ exists $\forall k$. Let there exist constants $c_k, 1 \leq k \leq K$, such that

(condition 2)

$$\lim_{n\to\infty} \frac{1}{Q^{(n)}_{11}/\pi_1} \sum_{r=1}^{n} \sum_{i,j\in\mathcal{A}_k} \pi_i (\rho(i)-c_k)(\rho(j)-c_k) {}_{\mathcal{H}}p^{(r)}_{ij} = 0 \quad \forall k,$$

and

(condition 3)

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{Q^{(n)}_{11}/\pi_1} \sum_{r=1}^{n} \sum_{i,j\in\mathcal{A}_k} \pi_i |\rho(i)\rho(j)| \; 1(|\rho(i)| > L, |\rho(j)| > L) {}_{\mathcal{H}}p^{(r)}_{ij} = 0 \quad \forall k.$$

Then,

$$\lim_{n\to\infty} \frac{var(\sum_{r=1}^{n} \rho_i)}{R^{(n)}_{11}/\pi_1} = \sum_{k=1}^{K} \pi^{\infty}_{\mathcal{A}_k} (\mu - c_k)^2.$$

Moreover, if $\pi^{\infty}_{\mathcal{A}_k}(c_k - \mu) \neq 0$ for some k, then $H_\rho = H$.

Remark. If $c_k = c_l$ for a pair of subsets $\mathcal{A}_k, \mathcal{A}_l$, then condition 1 is not needed for this particular pair.

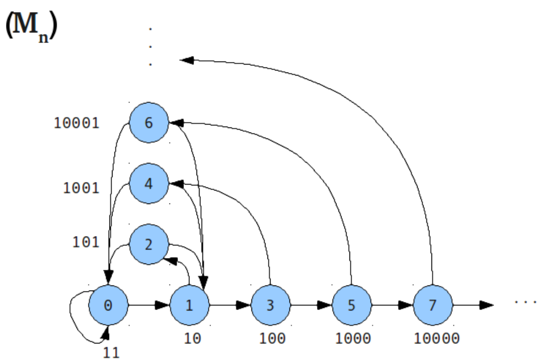
VI. Proofs

A. Proof of theorem II.3

$\rho_n := 1(Q_2(n) = 0)$ is an L_2 function of the chain $(X_1(n), X_2(n), Q_1(n), Q_2(n))$. We will use theorem V.2 with a single partition $\mathcal{A}_1 = \mathbb{N}$. Take $c_1 = 0$. $\mathcal{H} = \{X_1(n), X_2(n), Q_1(n), Q_2(n) \leq R\}$. Condition 3 holds trivially for bounded functions. Thus we are left with having to check the condition

$$\lim_{n\to\infty} \frac{1}{Q^{(n)}_{11}/\pi_1} \sum_{i,j:Q_{2,j}=0,Q_{2,i}=0} \pi_i \sum_{r=1}^{n} {}_{\mathcal{H}}p^{(r)}_{ij} = 0.$$

To see why this is true, note that $\sum_{i,j:Q_{2,j}=0 Q_{2,i}=0} \pi_i \sum_{r=1}^{\infty} {}_{\mathcal{H}}p^{(r)}_{ij}$ is bounded above by 1 plus the stationary time spent in the states $\{Q_2 = 0\}$ before the chain visits $\mathcal{H}$. Note that the length of an idle period for Q_2 has finite expectation. Also note, if an idle period begins at time $n + 1$, this implies due to the LQF policy that $Q_1(n) \leq R$, $Q_2(n) \leq R$, $X_1(n) \leq R$, and $X_2(n) \leq R$. Thus between successive idle periods of Q_2, the chain must visit $\mathcal{H}$. The stationary expected time spent in $\{Q_2 = 0\}$ without visiting $\mathcal{H}$ is therefore finite. Since $Q^{(n)}_{11} \to \infty$ (3), the above limit holds. Using theorem V.2, we conclude



X_n	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0
M_n	-	0	1	3	4	1	3	5	7	8	0	0	1	3	5	6	1	3	5	7	9

Fig. 2. Construction of the Markov chain, with an example sequence showing the correspondence with X_n

that $1(Q_2(n) = 0)$ has the same Hurst index as the chain $(X_1(n), X_2(n), Q_1(n), Q_2(n))$.

B. Proof of theorem III.3

We construct the following Markov chain (M_n) from the renewal process (X_n) (fig. 2):

- $M_n \in \{0, 1, 2, 3, \ldots\}$.
- $\{M_n = 0\} = \{X_{n-1}^n = 11\}$.
- For $k \in \{1, 2, \ldots\}$
  - $\{M_n = 2k - 1\} =$ $\{X_n = 0$ and k zeros since last arrival $\}$,
  - $\{M_n = 2k\} =$ $\{X_n = 1$ and k zeros since last arrival in $X_n\}$.

We establish some notation:

$$\begin{aligned}
(X_n),\quad & \text{stationary renewal process,}\\
& \text{interval-arrival lengths having the law of } T+1;\\
f_T(k) &:= P(T = k);\\
F_T(k) &:= P(T \le k);\\
\rho_n(X_{-\infty}^n) &:= -\log P(X_n | X_{-\infty}^{n-1});\\
\eta &:= E[\log P(X_1 | X_{-\infty}^0)].
\end{aligned}$$

One can easily check $\rho_n = \rho(M_n)$, with

- $\rho(0) = -\log f_T(0)$,
- $\rho(2k-1) = -\log P(T > k-1 | T \ge k-1)$,
- $\rho(2k) = -\log P(T = k | T \ge k)$.

We verify:

Lemma VI.1. *ρ_n is an L_2 function of M_n.*

Proof: Let π_i be the stationary distribution of (M_n). Note that $\pi_i > 0 \implies \rho_i < \infty$. We want to prove

$$\sum \rho(i)^2 \pi_i < \infty.$$

Note that $\pi_{2k+1} = \pi_{2k-1} P(T > k | T \ge k)$, and $\pi_{2k} = \pi_{2k-1} P(T = k | T \ge k)$ for $k = 1, 2, \ldots$. This gives

$$\begin{aligned}
\sum \rho(i)^2 \pi_i = \pi_0 \rho(0)^2 + \pi_1 \rho(1)^2 \\
+ \sum_{k=1}^{\infty} \pi_{2k-1} P(T = k | T \ge k) \log^2 P(T = k | T \ge k) \\
+ \sum_{k=1}^{\infty} \pi_{2k-1} P(T > k | T \ge k) \log^2 P(T > k | T \ge k),
\end{aligned}$$

$$\pi_0 \rho(0)^2 = (\sum_{k=1}^{\infty} \pi_{2k}) f_T(0) \log^2 f_T(0),$$

$$\pi_1 \rho(1)^2 = (\sum_{k=1}^{\infty} \pi_{2k})(1 - f_T(0)) \log^2 (1 - f_T(0)).$$

Since the $p \log^2 p$ terms are bounded above by 1, $\sum \rho_i^2 \pi_i \le 4$. ■

Now, to apply theorem V.2 we partition the state space into 3 sets as follows: $\mathcal{A}_1 = \{i > 0 : i \text{ even}\}$, $\mathcal{A}_2 = \{0\} \cup \{i \text{ odd} : \rho(i) \le -\log(1 - \epsilon_i)\}$, and $\mathcal{A}_3 = \{i \text{ odd} : \rho(i) > -\log(1 - \epsilon_i)\}$. Here we will will chose $\epsilon_i \downarrow 0$ later. Take $c_1 = c_2 = c_3 = 0$ and $\mathcal{H} = 1$ in that theorem. By the remark to the theorem, we don't need condition 1. We will check conditions 2 and 3 of theorem V.2 for each of the sets.

When $i, j \in \mathcal{A}_1$ notice ${}_1p_{ij}^{(r)} = 0$, so both conditions hold automatically. For $i, j \in \mathcal{A}_2$, condition 2 holds due to remark no. 2 because the limit of $\rho(i)$ as $i \to \infty$ is zero, and condition 3 holds because ρ is bounded on this set. Thus we focus on $i, j \in \mathcal{A}_3$. Define $\rho(i) =: -\log(1 - \tilde{\epsilon}_i)$. Let subsequence $\{i_k\} = \mathcal{A}_3$. We have $\tilde{\epsilon}_{i_k} \ge \epsilon_{i_k}$. $\pi_{i_k} \le \pi_1 \prod_{l=1}^{k} (1 - \tilde{\epsilon}_{i_k})$, and $\sum_1^{\infty} {}_1p_{i_k i_j}^{(r)} = \pi_{i_j} / \pi_{i_k}$. We have

$$\sum_i \rho(i) \pi_i \sum_j \rho(j) \sum_{r=1}^{n} {}_1p_{ij}^{(r)}$$

$$\le \sum_k \prod_{l=1}^{k} (1 - \tilde{\epsilon}_{i_l})(-\log(1 - \tilde{\epsilon}_{i_k})) \sum_{j>k} -\log(1 - \tilde{\epsilon}_{i_j}) \prod_{l=k+1}^{j} (1 - \tilde{\epsilon}_{i_l})$$

$$= \sum_j \sum_{k<j} (1 - \tilde{\epsilon}_{i_k}) \log(1 - \tilde{\epsilon}_{i_k})(1 - \tilde{\epsilon}_{i_j}) \log(1 - \tilde{\epsilon}_{i_j}) \prod_{l=1, l \ne k, j}^{j} (1 - \tilde{\epsilon}_{i_l})$$

$$< \sum_j j \prod_{l=3}^{j} (1 - \tilde{\epsilon}_{i_l}).$$

We can easily choose $\epsilon_i \downarrow 0$ such that this is finite. Dividing by $Q_{11}^{(n)}$, both conditions in theorem V.2 will be satisfied.

C. Proof of theorem IV.1

It can be verified that (also see the calculations in Mandelbrot's original paper [4]) X_n changes as follows: jumps by $E[T]$ on the first good day. Jumps by $-E[T]$ on the first bad day. Increases by $E[T | T \ge t] - E[T | T \ge t - 1]$ on the t^{th} good day ($t \ge 2$). Decreases by $E[T | T \ge t] - E[T | T \ge t - 1]$ on the t^{th} bad day. The first neutral day following t good days decreases X_n by $E[T|T \geq t] - t$. The first neutral day following t bad days increases X_n by $E[T|T \geq t] - t$.

Let $J_n = 1$(there is a transition at time n). Let $T_n := \inf_t\{t \geq 0 : W_{n-t-1} \neq W_{n-t-2}\}$ be the number of days since the last transition (0 on the first day following).

Then $M_n = (W_n, J_n, T_n)$ is a countable state, long range dependent Markov chain, with Hurst index $\frac{1}{2}(3-\alpha)$. Moreover, $\rho_n = |r_n|^d$ is a function of M_n:

- $\rho(\{g,b\}, 0, t) = (E[T|T \geq t+2] - E[T|T \geq t+1])^d$;
- $\rho(\{n\}, 0, \cdot) = 0$;
- $\rho(\{g,b\}, 1, \cdot) = (E[T])^d$;
- $\rho(\{n\}, 1, t) = (E[T|T \geq t+1] - (t+1))^d$.

Lemma VI.2.

$$E[T|T \geq t+2] - E[T|T \geq t+1] \to \frac{\alpha}{\alpha-1}, \qquad t \to \infty.$$

Proof:

$$P(T \geq s|T \geq t) = \frac{s^{-\alpha}}{t^{-\alpha}}, \qquad s \geq t;$$

$$\begin{aligned} E[T|T \geq& t+1] - E[T|T \geq t] = \\ &\sum_{s=t+1}^{\infty} P(T \geq s|T \geq t+1) - P(T \geq s|T \geq t) \\ &= ((t+1)^{\alpha} - t^{\alpha}) \sum_{s=t+1}^{\infty} s^{-\alpha} \to \frac{\alpha}{\alpha-1}, \end{aligned}$$

since $((t+1)^{\alpha} - t^{\alpha})/t^{\alpha-1} \to \alpha$ and $\frac{1}{\alpha-1}(t+2)^{-\alpha+1} = \int_{t+2}^{\infty} s^{-\alpha} ds < \sum_{s=t+1}^{\infty} s^{-\alpha} < \int_{t+1}^{\infty} s^{-\alpha} ds = \frac{1}{\alpha-1}(t+1)^{-\alpha+1}$. ∎

Lemma VI.3.

$$E[T|T \geq t] - t \leq \frac{t}{\alpha-1}.$$

Proof:

$$E[T|T \geq t] - t = \sum_{s=t}^{\infty} \frac{s^{-\alpha}}{t^{-\alpha}} \leq \int_t^{\infty} s^{-\alpha} ds = \frac{t}{\alpha-1}.$$

∎

We will utilize theorem V.2 with $\mathcal{A}_1 = (\{g,b\}, 0, \cdot), \mathcal{A}_2 = (\{n\}, 0, \cdot), \mathcal{A}_3 = (\{g,b\}, 1, \cdot), \mathcal{A}_4 = (\{n\}, 1, \cdot)$. $c_1 = c_4 = \left(\frac{\alpha}{\alpha-1}\right)^d, c_2 = c_3 = 0$. $\mathcal{H} = (\cdot, \cdot, 0)$. We have

$$\begin{aligned} \mathrm{var}(\rho_0) &\leq E\rho_0^2 = \sum_i \pi_i \rho(i)^2 \\ &= \sum_{i \notin \mathcal{A}_4} \pi_i \rho(i)^2 + \sum_{i \in \mathcal{A}_4} \pi_i \rho(i)^2 \\ &\leq C + \sum_{t=1}^{\infty} \frac{1}{2} P(T=t) (\frac{t}{\alpha-1})^{2d} < \infty, \end{aligned}$$

by lemma VI.3. As $\rho(i)$ is bounded when $i \notin \mathcal{A}_4$, the contribution to the sum is a constant C. We also used the fact that if $i = (\{n\}, 1, t-1)$, then $\pi_i = P(W_{-t} = n)P(T = t) = \frac{1}{2}P(T=t)$.

We need to first show that condition 1 holds:

$$\lim_{n\to\infty} \frac{1}{Q_{11}^{(n)}} \sum_{r=1}^{n} \sum_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} \pi_i |\rho(i) - \mu| |\rho(j) - \mu| {}_{\mathcal{H}}p_{ij}^{(r)} \to 0 \quad \forall k \neq l.$$

By inspection, the following transitions require visiting $\mathcal{H}$: (k,l) or $(l,k) = (1,2), (1,3), (2,4), (3,4)$. The sum is zero for these pairs. For (k,l) or $(l,k) = (1,4), (2,3)$, the condition is not needed due to the remark to theorem V.2.

Condition 2 reads

$$\lim_{n\to\infty} \frac{1}{Q_{11}^{(n)}} \sum_{i,j \in \mathcal{A}_k} \pi_i (\rho(i) - c_k)(\rho(j) - c_k) \sum_{r=1}^{n} {}_{\mathcal{H}}p_{ij}^{(r)} = 0 \quad \forall k.$$

For $k = 3, 4$, ${}_{\mathcal{H}}p_{ij}^{(r)} = 0$ because these states must go to $\mathcal{H}$ in one step. For $k = 1, 2$, we have chosen c_k such that $(\rho(i) - c_k) \to 0$ by lemma VI.2. The condition holds by remark no. 2.

Condition 3 also holds for $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$ because ρ is bounded on these sets. On $\mathcal{A}_4$, it holds because ${}_{\mathcal{H}}p_{ij}^{(r)} = 0$ as argued earlier. We finally have the conclusion:

$$\lim_{n\to\infty} \frac{\sum_{r=1}^{n} \mathrm{cov}(|r_0|^d, |r_n|^d)}{Q_{11}^{(n)}/\pi_1} = \sum_{k=1}^{K} \pi_k^{\infty} (\mu - c_k)^2 > 0.$$

ACKNOWLEDGEMENTS

Research support from the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-Scale Networks", from the NSF grant CNS-0910702, from the NSF Science & Technology Center grant CCF-0939370, "Science of Information", from Marvell Semiconductor Inc., and from the U.C. Discovery program is gratefully acknowledged.

REFERENCES

[1] G. Samorodnitsky and M. Taqqu, *Stable non-Gaussian processes: Stochastic models with infinite variance*. Chapman & Hall, 1994.

[2] P. Doukhan, G. Oppenheim, and M. Taqqu, *Theory and applications of long-range dependence*. Birkhauser, 2003.

[3] J. Beran, *Statistics for long-memory processes*. Chapman & Hall/CRC, 1994, vol. 61.

[4] B. Mandelbrot, "Forecasts of future prices, unbiased markets, and martingale models," *Journal of Business*, vol. 39, no. 1, pp. 242–255, 1966.

[5] M. Henry and P. Zaffaroni, "The long range dependence paradigm for macroeconomics and finance," *Theory and Applications of Long-Range Dependence*, pp. 417–438, 2003.

[6] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *Networking, IEEE/ACM Transactions on*, vol. 2, no. 1, pp. 1–15, 1994.

[7] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *Networking, IEEE/ACM Transactions on*, vol. 5, no. 6, pp. 835–846, 1997.

[8] K. Park and W. Willinger, "Self-similar network traffic: An overview," *Self-Similar Network Traffic and Performance Evaluation*, pp. 1–38, 2000.

[9] S. Resnick and G. Samorodnitsky, "Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues," *Queueing Systems*, vol. 33, no. 1, pp. 43–71, 1999.

[10] N. Likhanov, B. Tsybakov, and N. Georganas, "Analysis of an atm buffer with self-similar (fractal) input traffic," in *INFOCOM'95. Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. Proceedings. IEEE.* IEEE, 1995, pp. 985–992.

[11] N. Likhanov and R. Mazumdar, "Cell loss asymptotics for buffers fed with a large number of independent stationary sources," *Journal of Applied Probability*, vol. 36, no. 1, pp. 86–96, 1999.

[12] D. Heath, S. Resnick, and G. Samorodnitsky, "Patterns of buffer overflow in a class of queues with long memory in the input stream," *The Annals of Applied Probability*, vol. 7, no. 4, pp. 1021–1057, 1997.

[13] B. Zwart, S. Borst, and M. Mandjes, "Exact queueing asymptotics for multiple heavy-tailed on-off flows," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1. IEEE, 2001, pp. 279–288.

[14] A. Es-Saghouani and M. Mandjes, "On the correlation structure of a lévy-driven queue," *Journal of Applied Probability*, vol. 45, no. 4, pp. 940–952, 2008.

[15] V. Anantharam, "Scheduling strategies and long-range dependence," *Queueing systems*, vol. 33, no. 1, pp. 73–89, 1999.

[16] K. Park, G. Kim, and M. Crovella, "On the relationship between file sizes, transport protocols, and self-similar network traffic," in *Network Protocols, 1996. Proceedings., 1996 International Conference on*. IEEE, 1996, pp. 171–180.

[17] Z. Sahinoglu and S. Tekinay, "On multimedia networks: self-similar traffic and network performance," *Communications Magazine, IEEE*, vol. 37, no. 1, pp. 48–52, 1999.

[18] M. Markakis, E. Modiano, and J. Tsitsiklis, "Scheduling policies for single-hop networks with heavy-tailed traffic," *Proceedings of the 47th annual Allerton conference on Communication, control, and computing*, pp. 112–120, 2009.

[19] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1566–1579, 1995.

[20] M. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 4, pp. 269–280, 1994.

[21] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems," *Conference on Local Computer Networks: Proceedings*, p. 397, 1995.

[22] F. Fitzek and M. Reisslein, "Mpeg4 and h. 263 video traces for network performance," *IEEE Network*, vol. 15, no. 6, pp. 40–54, 2001.

[23] B. Oğuz and V. Anantharam, "Hurst index of functions of long range dependent Markov chains," *Journal of Applied Probability*, vol. 49, no. 2, 2012.

[24] K. Carpio and D. Daley, "Long-range dependence of Markov chains in discrete time on countable state space," *Journal of Applied Probability*, vol. 44, no. 4, pp. 1047–1055, 2007.

[25] N. Bingham, C. Goldie, and J. Teugels, *Regular Variation*. Cambridge University Press, 1989.

[26] B. Oğuz and V. Anantharam, "Compressing a long range dependent renewal process," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1443–1447.

[27] A. Barron, "Logically smooth density estimation." Ph.D. dissertation, Stanford University, Department of Electrical Engineering, 1985.

[28] I. Kontoyiannis, "Second-order noiseless source coding theorems," *Information Theory, IEEE Transactions on*, vol. 43, no. 4, pp. 1339–1341, 1997.

[29] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, no. 2, pp. 223–236, 2001.

[30] L. Giraitis, P. Robinson, and D. Surgailis, "A model for long memory conditional heteroscedasticity," *Annals of Applied Probability*, vol. 10, no. 3, pp. 1002–1024, 2000.

[31] K. Chung, *Markov chains with Stationary Transition Probabilities*. Berlin: Springer-Verlag, 1967.