fractal and fractional

Article

A Finite-State Stationary Process with Long-Range Dependence and Fractional Multinomial Distribution

Jeonghwa Lee

Department of Statistics, Truman State University, Kirksville, MO 63501, USA; jlee@truman.edu

Abstract: We propose a discrete-time, finite-state stationary process that can possess long-range dependence. Among the interesting features of this process is that each state can have different long-term dependency, i.e., the indicator sequence can have a different Hurst index for different states. Furthermore, inter-arrival time for each state follows heavy tail distribution, with different states showing different tail behavior. A possible application of this process is to model over-dispersed multinomial distribution. In particular, we define a fractional multinomial distribution from our model.

Keywords: long-range dependence; Hurst index; over-dispersed multinomial distribution

Citation: Lee, J. A Finite-State Stationary Process with Long-Range Dependence and Fractional Multinomial Distribution. *Fractal Fract.* **2022**, *6*, 596. https://doi.org/10.3390/fractalfract6100596

Academic Editors: Libo Feng, Yang Liu and Lin Liu

Received: 17 September 2022
Accepted: 11 October 2022
Published: 14 October 2022

1. Introduction

Long-range dependence (LRD) refers to a phenomenon where correlation decays slowly with the time lag in a stationary process in a way that the correlation function is no longer summable. This phenomenon was first observed by Hurst [1,2] and since then it has been observed in many fields such as economics, hydrology, internet traffic, queueing networks, etc. [3–6]. In a second order stationary process, LRD can be measured by the Hurst index H [7,8],

$$H = \inf\{h : \limsup_{n\to\infty} n^{-2h+1} \sum_{k=1}^{n} cov(X_1, X_k) < \infty\}.$$

Note that $H \in (0,1)$, and if $H \in (1/2,1)$, the process possesses a long-memory property.

Among the well-known stochastic processes that are stationary and possess long-range dependence are fractional Gaussian noise (FGN) [9] and fractional autoregressive integrated moving average processes (FARIMA) [10,11].

Fractional Gaussian noise X_j is a mean-zero, stationary Gaussian process with covariance function:

$$\gamma(j) := cov(X_0, X_j) = \frac{var(X_0)}{2}(|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H})$$

where $H \in (0,1)$ is the Hurst parameter. The covariance function obeys the power law with exponent $2H-2$ for large lag,

$$\gamma(j) \sim var(X_0)H(2H-1)j^{2H-2} \text{ as } j \to \infty.$$

If $H \in (1/2,1)$, then the covariance function decreases slowly with the power law, and $\sum_j \gamma(j) = \infty$, i.e., it has the long-memory property.

A FARIMA(p,d,q) process $\{X_t\}$ is the solution of:

$$\phi(B)\nabla^d X_t = \theta(B)\epsilon_t,$$

where p, q are positive integers, d is real, B is the backward shift, $BX_t = X_{t-1}$, and the fractional-differencing operator ∇^d, autoregressive operator ϕ, and moving average operator θ are, respectively,

$$\nabla^d = (1-B)^d = \sum_{k=1}^{\infty} \frac{d(d-1)\cdots(d+1-k)}{k!}(-B)^k,$$
$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p,$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q.$$

where $\{\epsilon_t\}$ is the white-noise process, which consists of iid random variables with the finite second moment. Here, the parameter d manages the long-term dependence structure, and by its relation to the Hurst index, $H = d + 1/2$, $d \in (0, 1/2)$ corresponds to the long-range dependence in the FARIMA process.

Another class of stationary processes that can possess long-range dependence is from the countable-state Markov process [12]. In a stationary, positive recurrent, irreducible, aperiodic Markov chain, the indicator sequence of visits to a certain state is long-range dependent if and only if return time to the state has an infinite second moment, and this is possible only when the Markov chain has infinite state space. Moreover, if one state has the infinite second moment of return time, then all the other states also have the infinite second moment of return time, and all the states have the same rate of dependency; that is, the indicator sequence of each state is long-range dependence with the same Hurst index.

In this paper, we develop a discrete-time finite-state stationary process that can possess long-range dependence. We define a stationary process $\{X_i, i \in \mathbb{N}\}$ where the number of possible outcomes of X_i is finite, $S = \{0, 1, \cdots, m\}$ for any $m \in \mathbb{N}$, and for $k = 1, 2, \cdots, m$,

$$cov(I_{\{X_i=k\}}, I_{\{X_j=k\}}) = c_k'|i-j|^{2H_k-2}, \quad (1)$$

for any $i, j \in \mathbb{N}, i \neq j$, and some constants $c_k' \in \mathbb{R}_+$, $H_k \in (0,1)$. This leads to:

$$cov(X_i, X_j) \sim c_{k'}'|i-j|^{2H_{k'}-2} \quad \text{as } |i-j| \to \infty, \quad (2)$$

where $k' = argmax_k\{H_k; k = 1, \cdots, m\}$. If $H_{k'} = \max\{H_k; k = 1, \cdots, m\} \in (1/2, 1)$, (1.2) implies that as $n \to \infty$, $\sum_{i=1}^n cov(X_1, X_i)$ diverges with the rate of $|n|^{2H_{k'}-1}$, and the process is said to have long-memory with Hurst parameter $H_{k'}$. Furthermore, from (1.1), for $k = \{1, \cdots, m\}$, the process $\{I_{\{X_i=k\}}; i = 1, 2, \cdots\}$ is long-range dependence if $H_k \in (1/2, 1)$. In particular, if $H_i \neq H_j$, then the states "i" and "j" produce different levels of dependence. For example, if $H_i < 1/2 < H_j$, then the state "j" produces a long-memory counting process whereas state "i" produces a short-memory process.

A possible application of our stochastic process is to model the over-dispersed multinomial distribution. In the multinomial distribution, there are n trials, each trial results in one of the finite outcomes, and the outcomes of the trials are independent and identically distributed. When applying the multinomial model to real data, it is often observed that the variance is larger than what it is assumed to be, which is called over-dispersion, due to the violation of the assumption that trials are independent and have identical distribution [13,14], and there have been several ways to model an overdispersed multinomial distribution [15–18].

Our stochastic process provides a new method to model an over-dispersed multinomial distribution by introducing dependency among trials. In particular, the variance of the number of a certain outcomes among n trials is asymptotically proportional to the fractional exponent of n, from which we define:

$$Y_k := \sum_{i=1}^{n} I_{\{X_i=k\}} \text{ for } k = 1, 2, \cdots, m,$$

and call the distribution of $(Y_1, Y_2, \cdots, Y_m)$ the fractional multinomial distribution.

The work in this paper is an extension of the earlier work of the generalized Bernoulli process [19], and the process in this paper is reduced to the generalized Bernoulli process if there are only two states in the possible outcomes of X_i, e.g., $S = \{0, 1\}$.

In Section 2, a finite state stationary process that can possess long-range dependence is developed. In Section 3, the properties of our model are investigated with regard to tail behavior and moments of inter-arrival time of a certain state "k", and conditional probability of observing a state "k" given the past observations in the process. In Section 4, the fractional multinomial distribution is defined, followed by the conclusions in Section 5. Some proofs of propositions and theorems are in Section 6.

Throughout this paper, $\{i, i_0, i_1, \cdots\}, \{i', i'_0, i'_1, \cdots\} \subset \mathbb{N}$, with $i_0 < i_1 < i_2 < \cdots$, and $i'_0 < i'_1 < i'_2 < \cdots$. For any set $A = \{i_0, i_1, \cdots, i_n\}$, $|A| = n + 1$, the number of elements in the set A, and for the empty set, we define $|\emptyset| = 0$.

2. Finite-State Stationary Process with Long-Range Dependence

We define the stationary process $\{X_i, i \in \mathbb{N}\}$ where the set of possible outcomes of X_i is finite, $S = \{0, 1, \cdots, m\}$, for $m \in \mathbb{N}$, with the probability that we observe a state "k" at time i is $P(X_i = k) = p_k > 0$, for $k = 0, 1, \cdots, m$, and $\sum_{k=0}^{m} p_k = 1$.

For any set $A = \{i_0, i_1, \cdots, i_n\} \subset \mathbb{N}$, define the operator:

$$L^*_{H,p,c}(A) := p \prod_{j=1,\cdots,n} (p + c|i_j - i_{j-1}|^{2H-2}).$$

If $A = \emptyset$, define $L^*_{H,p,c}(A) := 1$, and if $A = \{i_0\}$, $L^*_{H,p,c}(A) := p$.

Let $\mathbf{H} = (H_1, H_2, \cdots, H_m)$, $\mathbf{p} = (p_1, p_2, \cdots, p_m)$, $\mathbf{c} = (c_1, c_2, \cdots, c_m)$ be vectors of length m, and $\mathbf{H}, \mathbf{p}, \mathbf{c} \in (0, 1)^m$. We are now ready to define the following operators.

Definition 1. *Let $A_0, A_1, \cdots, A_m \subset \mathbb{N}$ be pairwise disjoint, and $A_0 = n' > 0$. Define,*

$$L^*_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, A_2, \cdots, A_m) := \prod_{k=1,\cdots m} L^*_{H_k,p_k,c_k}(A_k),$$

and,

$$D^*_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, A_2, \cdots, A_m; A_0) := \sum_{\ell=0}^{n'} (-1)^\ell \sum_{\substack{|B|=\ell \\ B \subset A_0}} \sum_{\substack{B_i \subset B \\ B_i \cap B_j = \emptyset \\ \cup B_i = B}} L^*_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1 \cup B_1, A_2 \cup B_2, \cdots, A_m \cup B_m).$$

For ease of notation, we denote $D^*_{\mathbf{H},\mathbf{p},\mathbf{c}}$, $L^*_{\mathbf{H},\mathbf{p},\mathbf{c}}$, and $L^*_{H_k,p_k,c_k}$ by $\mathbf{D}^*, \mathbf{L}^*, L^*_k$, respectively. Note that if $A_0 = \{i_0\}$,

$$\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) = \prod_{k=1,\cdots,m} L^*_k(A_k) \left(1 - \sum_{k'=1}^{m} \frac{L^*_{k'}(A_{k'} \cup \{i_0\})}{L^*_{k'}(A_{k'})}\right). \tag{3}$$

For any pairwise disjoint sets $A_0, A_1, \cdots A_m \subset \mathbb{N}$, if $\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) > 0$, then $\{X_i; i \in \mathbb{N}\}$ is well defined stationary process with the following probabilities:

$$P(\cap_{i \in A_k}\{X_i = k\}) = L^*_k(A_k), \text{ for } k = 1, \cdots, m, \tag{4}$$

$$P(\cap_{k=1,\cdots,m} \cap_{i \in A_k} \{X_i = k\}) = \prod_{k=1,\cdots,m} L^*_k(A_k), \tag{5}$$

$$P(\cap_{k=0,\cdots,m} \cap_{i \in A_k} \{X_i = k\}) = \mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0). \tag{6}$$

In particular, if the stationary process with the probability above is well defined, then, for $k, k' = 1, \cdots, m$, we have:

$$
\begin{aligned}
P(X_i = k, X_j = k) &= p_k(p_k + c_k|j-i|^{2H_k-2}), \\
P(X_i = k, X_j = k') &= p_k p_{k'},
\end{aligned}
$$

$$
\begin{aligned}
P(X_i = 0, X_j = 0) &= 1 - 2\sum_{k=1,\cdots,m} P(X_i = k) + \sum_{k,k'=1,\cdots,m} P(X_i = k, X_j = k') \\
&= 1 - 2\sum_{k=1}^{m} p_k + \sum_{k=1}^{m} p_k(p_1 + p_2 + \cdots + p_m + c_k|i-j|^{2H_k-2}) \\
&= p_0^2 + \sum_{k=1}^{m} p_k c_k |i-j|^{2H_k-2}, \\
P(X_i = k, X_j = 0) &= P(X_i = 0, X_j = k) = p_k(1 - p_1 - p_2 - \cdots - p_m - c_k|i-j|^{2H_k-2}) \\
&= p_k(p_0 - c_k|i-j|^{2H_k-2}).
\end{aligned}
$$

As a result, for $i \neq j, i, j \in \mathbb{N}, k \neq k', k, k' \in \{1, 2, \cdots, m\}$,

$$cov(I_{\{X_i=k\}}, I_{\{X_j=k\}}) = p_k c_k |i-j|^{2H_k-2}, \tag{7}$$

$$cov(I_{\{X_i=k\}}, I_{\{X_j=k'\}}) = 0, \tag{8}$$

$$cov(I_{\{X_i=0\}}, I_{\{X_j=0\}}) = \sum_{k=1}^{m} p_k c_k |i-j|^{2H_k-2}, \tag{9}$$

$$cov(I_{\{X_i=k\}}, I_{\{X_j=0\}}) = -p_k c_k |i-j|^{2H_k-2}. \tag{10}$$

Note that $(\{I_{\{X_i=1\}}\}_{i\in\mathbb{N}}, \{I_{\{X_i=2\}}\}_{i\in\mathbb{N}}, \cdots, \{I_{\{X_i=m\}}\}_{i\in\mathbb{N}})$ are m generalized Bernoulli processes with Hurst parameter, $H_1, H_2, \cdots, H_m$, respectively (see [19]). However, they are not independent, since for $\ell \neq k, \ell \in \{1, 2, \cdots, m\}$,

$$P(\{I_{\{X_i=\ell\}} = 1\} \cap \{I_{\{X_i=k\}} = 1\}) = 0 \neq P(I_{\{X_i=\ell\}} = 1)P(I_{\{X_i=k\}} = 1) = p_\ell p_k.$$

Further, we have,

$$
\begin{aligned}
cov(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\
&= \sum_{k,k'} kk' P(I_{\{X_i=k\}} = 1, I_{\{X_j=k'\}} = 1) - \sum_{k,k'} kk' p_k p_{k'} \\
&= \sum_{k=1,\cdots,m} k^2 p_k c_k |i-j|^{2H_k-2}.
\end{aligned}
$$

Therefore, the process $\{X_i\}_{i\in\mathbb{N}}$ possesses long-range dependence if $\min\{H_1, \cdots, H_k\} > 1/2$.

All the results that appear in this paper are valid regardless of how the finite-state space of X_i is defined. More specifically, given that: $\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) > 0$ for any pairwise disjoint sets $A_0, A_1, \cdots A_m \subset \mathbb{N}$, we can define probability (4)–(6) with any state space $S = \{s_0, s_1, s_2, \cdots, s_m\} \subset \mathbb{R}$ for any $m \in \mathbb{N}$ in the following way.

$$
\begin{aligned}
&P(\cap_{i\in A_k}\{X_i = s_k\}) = L_k^*(A_k), \text{ for } k = 1, \cdots, m, \\
&P(\cap_{k=1,\cdots,m} \cap_{i\in A_k} \{X_i = s_k\}) = \prod_{k=1,\cdots,m} L_k^*(A_k), \\
&P(\cap_{k=0,\cdots,m} \cap_{i\in A_k} \{X_i = s_k\}) = \mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0).
\end{aligned}
$$

Note that the only difference is that the space "k" is replaced by "s_k". As a result, we can obtain the same results as (7)–(10), except that $I_{\{X_i=k\}}$ is replaced by $I_{\{X_i=s_k\}}$, and we get:

$$\begin{aligned} cov(X_i, X_j) &= cov(X_i - s_0, X_j - s_0) \\ &= \sum_{k,k'=1,\cdots,m} s_k s'_k P(I_{\{X_i=s_k\}} = 1, I_{\{X_j=s'_k\}} = 1) - \sum_{k,k'=1,\cdots,m} s_k s'_k p_k p_{k'} \\ &= \sum_{k=1,\cdots,m} (s_k - s_0)^2 p_k c_k |i-j|^{2H_k-2}. \end{aligned}$$

In a similar way, all the results in this paper can be easily transfered to any finite-state space $S \subset \mathbb{R}$. For the sake of simplicity, we assume $S = \{0, 1, \cdots, m\}, m \in \mathbb{N}$, without loss of generality, and define $S^0 := \{1, \cdots, m\}$.

Now, we will give a restriction on the parameter values, $\{H_k, p_k, c_k; k \in S^0\}$, which will make $\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) > 0$ for any pairwise disjoint sets $A_0, \cdots A_m \subset \mathbb{N}$; therefore, the process $\{X_i\}$ is well-defined with the probability (4)–(6).

ASSUMPTIONS:

(A.1) $c_k, H_k, p_k \in (0, 1)$ for $k \in S^0$.

(A.2) For any $i_0 < i_1 < i_2, i_0, i_1, i_2 \in \mathbb{N}$,

$$\sum_{k=1}^{m} \frac{(p_k + c_k|i_1 - i_0|^{2H_k-2})(p_k + c_k|i_2 - i_1|^{2H_k-2})}{p_k + c_k|i_2 - i_0|^{2H_k-2}} < 1. \tag{11}$$

For the rest of the paper, it is assumed that ASSUMPTIONS (A.1, A.2) hold.

Remark 1. *(a). (11) holds if,*

$$\sum_{k=1}^{m} \frac{(p_k + c_k)(p_k + c_k)}{p_k + c_k 2^{2H_k-2}} < 1,$$

since,

$$\frac{(p_k + c_k|i_1 - i_0|^{2H_k-2})(p_k + c_k|i_2 - i_1|^{2H_k-2})}{(p_k + c_k|i_2 - i_0|^{2H_k-2})}$$

is maximized when $i_2 - i_0 = 2, i_1 - i_0 = 1$, as it was seen in Lemma 2.1 of [19].

(b). If $(i_1 - i_0)/(i_2 - i_0) \to 0, (i_2 - i_1)/(i_2 - i_0) \to 1$ with $i_2 - i_0 \to \infty$ in (11), then we have:

$$\sum_{k=1}^{m} p_k + c_k|i_1 - i_0|^{2H_k-2} < 1, \tag{12}$$

and this, together with (11), implies that for any set $\{A_k, i'_k\} \subset \mathbb{N}$,

$$\sum_{k=1}^{m} \frac{L_k^*(A_k \cup \{i'_k\})}{L_k^*(A_k)} < 1.$$

This means that for any $A_0 = \{i_0\} \subset \mathbb{N}$, $\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) > 0$ by (3).

(c). From (12), $\sum_{k=1}^{m} c_k < 1 - \sum_{k=1}^{m} p_k = p_0$.

(d). If $m = 1$, (11) is reduced to (2.7) in the Lemma 2.1 in [19].

Now we are ready to show that $\{X_i, i \in \mathbb{N}\}$ is well defined with probability (4)–(6).

Proposition 1. *For any disjoint sets $A_0, A_1, A_2, \cdots, A_m \subset \mathbb{N}, A_0 \neq \emptyset$,*

$$\mathbf{D}^*(A_1, A_2, \cdots, A_m; A_0) > 0.$$

The next theorem shows that the stochastic process $\{X_i, i \in \mathbb{N}\}$ defined with probability (4)–(6) is stationary, and it has long-range dependence if $\max\{H_k, k \in S^0\} > 1/2$. Furthermore, the indicator sequence of each state is stationary, and has long-range dependence if its Hurst exponent is greater than $1/2$.

Theorem 1. *$\{X_i, i \in \mathbb{N}\}$ is a stationary process with the following properties.*
i.
$$P(X_i = k) = p_k, \text{ for } k \in S^0.$$
ii.
$$cov(I_{\{X_i=k\}}, I_{\{X_j=k\}}) = p_k c_k |i-j|^{2H_k-2}, \text{ for } k \in S^0,$$
and
$$cov(I_{\{X_i=0\}}, I_{\{X_j=0\}}) \sim p_{k'} c_{k'} |i-j|^{2H_{k'}-2}, \text{ as } |i-j| \to \infty$$
where $k' = argmax_k H_k$.
iii.
$$cov(X_i, X_j) = \sum_{k=1}^{m} k^2 p_k c_k |i-j|^{2H_k-2}, \text{ for } i \neq j.$$

Proof. By Proposition 1, $\{X_i\}$ is a well-defined stationary process with probability (4)–(6). The other results follow by (7)–(10). □

3. Tail Behavior of Inter-Arrival Time and Other Properties

For $k \in S^0$, $\{I_{\{X_i=k\}}\}_{i\in\mathbb{N}}$ is a stationary process in which the event $\{X_i = k\}$ is recurrent, persistent, and aperiodic (here, we follow the terminology and definition in [20]). We define a random variable T^i_{kk} as the inter-arrival time between the i-th "k" from the previous "k", i.e.,
$$T^i_{kk} := \inf\{i > 0 : X_{i+T^{i-1}_{kk}} = k\},$$
with $T^0_{kk} := 0$. Since $\{I_{\{X_i=k\}}\}_{i\in\mathbb{N}}$ is GBP with parameters (H_k, p_k, c_k) for $k \in S^0$, $T^2_{kk}, T^3_{kk}, \cdots$ are iid (see page 9 [21]). Therefore, we will denote the inter-arrival time between two consecutive observations of k as T_{kk}. The next Lemma is directly obtained from Theorem 3.6 in [21].

Lemma 1. *For $k \in S^0$, the inter-arrival time for state k, T_{kk}, satisfies the following.*
i. T_{kk} has a mean of $1/p_k$. It has an infinite second moment if $H_k \in (1/2, 1)$.
ii.
$$P(T_{kk} > t) = t^{2H_k-3} L_k(t),$$
where L_k is a slowly varying function that depends on the parameter H_k, p_k, c_k.

The first result i in Lemma 1 is similar to Lemma 1 in [22]. However, here, we have a finite-state stationary process, whereas countable-state space Markov chain was assumed in [22]. Now, we investigate the conditional probabilities and the uniqueness of our process.

Theorem 2. *Let $A_0, A_1, \cdots, A_m$ be disjoint subsets of $\mathbb{N}$. For any $\ell \in S^0$ such that $\max A_\ell > \max A_0$, and for $i' \notin \cup_{k=0}^m A_k$ such that $i' > \max A_\ell$, the conditional probability satisfies the following:*
$$P(X_{i'} = \ell | \cap_{k=0,\cdots,m} \cap_{i\in A_k} \{X_i = k\}) = p_\ell + c_\ell |i' - \max A_\ell|^{2H_\ell - 2}.$$

If there has been no interruption of "0" after the last observation of "ℓ", then the chance to observe "ℓ" depends on the distance between the current time and the last time of observation of "ℓ", regardless of how other states appeared in the past. This can be considered as a generalized Markov property. Moreover, this chance to observe "ℓ" decreases as the distance increases, following the power law with exponent $2H_\ell - 2$.

Proof. The result follows from the fact that:

$$P(\{X_{i'} = \ell\} \cap_{\substack{i\in A_k \\ k\in S^0}} \{X_i = k\}) = P(\cap_{i\in A_k, k\in S^0}\{X_i = k\}) \times (p_\ell + c_\ell|i' - \max A_\ell|^{2H_\ell - 2}),$$

since there is no $i \in A_0$ between i' and $\max A_\ell$. □

In a countable state space Markov chain, long-range dependence is possible only when it has infinite state space, and additionally if it is stationary, positive recurrent, irreducible, aperiodic Markov chain, then each state should have the same long-term memory, i.e., sequence indicators have the same Hurst exponent for all states [22]. By relaxing the Markov property, long-range dependence was made possible in a finite-state stationary process, also with different Hurst parameter for different states.

Theorem 3. *Let $A_0, A_1, \cdots, A_m$ be disjoint subsets of $\mathbb{N}$. For $\ell \in S^0$ such that $\max A_\ell < \max A_0$, and $i'_1, i'_2, i'_3 \notin \cup_{k=0}^m A_k$ such that $i'_1, i'_2, i'_3 > \max A_0$, and $i'_2 > i'_3$, the conditional probability satisfies the following:*
a.

$$p_\ell + c_\ell|i'_1 - \max A_\ell|^{2H_\ell - 2} > P(X_{i'_1} = \ell| \cap_{i\in A_k, k\in S^0} \{X_i = k\}).$$

b.

$$\frac{P(X_{i'_2} = \ell| \cap_{i\in A_k, k\in S^0} \{X_i = k\})}{P(X_{i'_3} = \ell| \cap_{i\in A_k, k\in S^0} \{X_i = k\})} > \frac{p_\ell + c_\ell|i'_2 - \max A_\ell|^{2H_\ell - 2}}{p_\ell + c_\ell|i'_3 - \max A_\ell|^{2H_\ell - 2}}.$$

Theorem 4. *A stationary process with (4)–(6) is the unique stationary process that satisfies*
i. for $k \in S$:

$$P(X_i = k) = p_k, \quad \text{where } p_k > 0 \text{ and } \sum_{k=0}^{m} p_k = 1,$$

ii. for $k \in S^0$ and any $i, j \in \mathbb{N}, i \neq j$,

$$cov(I_{\{X_i = k\}}, I_{\{X_j = k\}}) = c'_k|i - j|^{2H_k - 2},$$

for some constants $c'_k \in \mathbb{R}_+, H_k \in (0, 1)$,
iii. for any sets, $A \subset S^0$ and $\{i_k; k \in A\} \subset \mathbb{N}$,

$$P(\cap_{k\in A}\{X_{i_k} = k\}) = \prod_{k\in A} p_k,$$

iv. for $\ell \in S^0$, there is a function $h_\ell(\cdot)$ such that,

$$P(X_{i'} = \ell| \cap_{i\in A_k, k\in S^0} \{X_i = k\}) = h_\ell(i' - \max A_\ell)$$

for disjoint subsets, $A_0, A_1, \cdots, A_m, \{i'\} \subset \mathbb{N}$, such that $A_\ell \neq \emptyset$, $i' > \max A_\ell$, and $\max A_\ell > \max A_0$ (A_0 can be the empty set).

Proof. Let X^* be a stationary process that satisfies *i*–*iv*. By *i*, *ii*,

$$P(X^*_{i_0} = k, X^*_{i_1} = k) = cov(I_{\{X^*_{i_0} = k\}}, I_{\{X^*_{i_1} = k\}}) + p_k^2 = c'_k|i_0 - i_1|^{2H_k - 2} + p_k^2,$$

which results in:

$$h_k(i_0 - i_1) = P(X^*_{i_1} = k|X^*_{i_0} = k) = p_k + (c'_k/p_k)|i_0 - i_1|^{2H_k - 2}.$$

Therefore, by *iv*,

$$\begin{aligned} P(X^*_{i_0} = k, X^*_{i_1} = k, X^*_{i_2} = k, \cdots, X^*_{i_n} = k) &= p_k \prod_{j=1}^{n} h_k(i_j - i_{j-1}) \\ &= L^*_k(\{i_0, i_2, \cdots, i_n\}), \end{aligned}$$

where $L^*_k = L^*_{H_k, p_k, c'_k/p_k}$. Furthermore, by applying *iii*, *iv* to X^*,

$$P(\cap_{i \in A_k, k \in S^0} \{X_i = k\}) = \prod_{k=1,\cdots,m} L^*_k(A_k).$$

This implies that X^* satisfies (4)–(6) with $c_k = c'_k/p_k$ for $k \in S^0$. □

4. Fractional Multinomial Distribution

In this section, we define a fractional multinomial distribution that can serve as an over-dispersed multinomial distribution.

Note that $\sum_{i=1}^{n} I_{\{X_i=k\}}$ has mean np_k for $k \in S$. Further, as $n \to \infty$,

$$var\Big(\sum_{i=1}^{n} I_{\{X_i=k\}}\Big) \sim \begin{cases} (p_k(1-p_k) + \dfrac{c'_k}{2H_k - 1})n & H_k \in (0, 1/2), \\ c'_k n \ln n & H_k = 1/2, \\ \dfrac{c'_k}{2H_k - 1}|n|^{2H_k}, & H_k \in (1/2, 1), \end{cases}$$

for $k \in S^0$, and,

$$var\Big(\sum_{i=1}^{n} I_{\{X_i=0\}}\Big) \sim \begin{cases} (p_{k'}(1-p_{k'}) + \dfrac{c'_{k'}}{2H_{k'} - 1})n & H_{k'} \in (0, 1/2), \\ c'_{k'} n \ln n & H_{k'} = 1/2, \\ \dfrac{c'_{k'}}{2H_{k'} - 1}|n|^{2H_{k'}}, & H_{k'} \in (1/2, 1), \end{cases}$$

where $k' = argmax_k\{H_k; k \in S^0\}$, and $c'_k = p_k c_k$. It also has the following covariance.

$$cov\Big(\sum_{i=1}^{n} I_{\{X_i=k\}}, \sum_{i=1}^{n} I_{\{X_i=k'\}}\Big) = -np_k p_{k'},$$

$$cov\Big(\sum_{i=1}^{n} I_{\{X_i=0\}}, \sum_{i=1}^{n} I_{\{X_i=k\}}\Big) = -np_0 p_k - \sum_{\substack{i \neq j \\ i,j=1,\cdots,n}} c'_k |i-j|^{2H_k - 2},$$

for $k, k' \in S^0$.

We define $Y_k := \sum_{i=1}^{n} I_{\{X_i=k\}}$, for $k \in S$, and a fixed n, and call its distribution fractional multinomial distribution with parameters $n, \mathbf{p}, \mathbf{H}, \mathbf{c}$.

If $\mathbf{c} = \mathbf{0}$, $(Y_0, Y_1, Y_2, \cdots, Y_m)$ follows a multinomial distribution with parameters $n, \mathbf{p}$, and $E(Y_k) = np_k, var(Y_k) = np_k(1-p_k), cov(Y_k, Y_{k'}) = -np_k p_{k'}$, for $k, k' \in S, k \neq k'$, and $p_0 = 1 - \sum_{i=1}^{m} p_i$.

If $\mathbf{c} \neq \mathbf{0}$, $(Y_0, Y_1, \cdots, Y_m)$ can serve as over-dispersed multinomial random variables with:

$$E(Y_k) = np_k, \quad Var(Y_k) = np_k(1-p_k)(1 + \psi_{n,k}),$$

where the over-dispersion parameter $\psi_{n,k}$ is as follows.

$$\psi_{n,k} \sim \begin{cases} \dfrac{c}{(1-p_k)(2H_k-1)} & \text{if } H_k \in (0,1/2), \\ \dfrac{c\ln n}{1-p_k} - 1 & \text{if } H_k = 1/2, \\ \dfrac{cn^{2H_k-1}}{(1-p_k)2H_k-1} - 1 & \text{if } H_k \in (1/2,1), \end{cases}$$

for $k \in S^0$, and,

$$\psi_{n,0} \sim \begin{cases} \dfrac{c}{(1-p_{k'})(2H_{k'}-1)} & \text{if } H_{k'} \in (0,1/2), \\ \dfrac{c\ln n}{1-p_{k'}} - 1 & \text{if } H_{k'} = 1/2, \\ \dfrac{cn^{2H_{k'}-1}}{(1-p_{k'})2H_{k'}-1} - 1 & \text{if } H_{k'} \in (1/2,1), \end{cases}$$

where $k' = argmax_k\{H_k; k \in S^0\}$, as $n \to \infty$. If $H_k \in (0,1/2)$, the over-dispersion parameter $\psi_{n,k}$ remains stable as n increases, whereas if $H_k \in (1/2,1)$ the over-dispersed parameter $\psi_{n,k}$ increases with the rate of fractional exponent of n, n^{2H_k-1}.

5. Conclusions

A new method for modeling long-range dependence in discrete-time finite-state stationary process was proposed. This model allows different states to have different Hurst indices except that for the base state "0", the Hurst exponent is the maximum Hurst index of all other states. Inter-arrival time for each state follows a heavy tail distribution, and its tail behavior is different for different states. The other interesting feature of this process is that the conditional probability to observe a state "k" (k is not the base state "0") depends on the Hurst index H_k and the time difference between the last observation of "k" and the current time, no matter how other states appeared in the past, given that there was no base state observed since the last observation of "k". From the stationary process developed in this paper, we defined a fractional multinomial distribution that can express a wide range of over-dispersed multinomial distributions; each state can have a different over-dispersion parameter that can behave as an asymptotically constant or grow with a fractional exponent of the number of trials.

6. Proofs

Lemma 2. *For any* $\{a_0, a_1, \cdots, a_n, a'_0, a'_1, \cdots, a'_n\} \subset \mathbb{R}_+$ *that satisfies* $a_0 - \sum_{i=1}^{j} a_i > 0, a'_0 - \sum_{i=1}^{j} a'_i > 0$ *for* $j = 1, 2, \cdots, n$,
i. if,

$$\frac{a_0}{a'_0} \geq \frac{a_1}{a'_1} \geq \cdots \geq \frac{a_n}{a'_n},$$

then,

$$\frac{a_0 - a_1 - a_2 - \cdots - a_n}{a'_0 - a'_1 - a'_2 - \cdots - a'_n} \geq \frac{a_0}{a'_0}.$$

ii. If,

$$\frac{a_0}{a'_0} < \frac{a_1}{a'_1} \leq \cdots \leq \frac{a_n}{a'_n},$$

then,

$$\frac{a_0 - a_1 - a_2 - \cdots - a_n}{a'_0 - a'_1 - a'_2 - \cdots - a'_n} \leq \frac{a_0}{a'_0}.$$

iii. For any $\{a_0, a_1, \cdots, a_n, a'_0, a'_1, \cdots, a'_n\} \subset \mathbb{R}_+$,

$$\max_i \frac{a_i}{a'_i} \geq \frac{a_1 + a_2 + \cdots + a_n}{a'_1 + a'_2 + \cdots + a'_n} \geq \min_i \frac{a_i}{a'_i}.$$

Proof. *i* and *ii* were proved in Lemma 5.2 in [19].
For *iii*, define b_j such that,

$$\frac{a_j}{a'_j} = b_j.$$

Then,

$$\frac{a_1 + a_2 + \cdots + a_n}{a'_1 + a'_2 + \cdots + a'_n} = \frac{b_1 a'_1 + b_2 a'_2 + \cdots + b_n a'_n}{a'_1 + a'_2 + \cdots + a'_n}$$

which is weighted average of $\{b_j, j = 1, \cdots, n\}$. □

To ease our notation, we will denote:

$$\mathbf{L}^*(A_1, A_2, \cdots, A_{k-1}, A_k \cup \{i\}, A_{k+1}, \cdots, A_m)$$

by,

$$\mathbf{L}^*(\cdots, A_k \cup \{i\}, \cdots),$$

and,

$$\mathbf{L}^*(\cdots, A_k \cup \{i\}, A_{k'} \cup \{j\}, \cdots) = \mathbf{L}^*(A^*_1, A^*_2, \cdots, A^*_m)$$

where, if $k \neq k'$,

$$A^*_i = \begin{cases} A_i \text{ if } i \neq k, k,' \\ A_i \cup \{i\} \text{ if } i = k, \\ A_i \cup \{j\} \text{ if } i = k', \end{cases}$$

and if $k = k'$,

$$A^*_i = \begin{cases} A_i \text{ if } i \neq k, \\ A_i \cup \{i \cup\} \text{ if } i = k. \end{cases}$$

$\mathbf{D}^*(\cdots, A_k \cup \{i\}, \cdots)$ and $\mathbf{D}^*(\cdots, A_k \cup \{i\}, A_{k'} \cup \{j\}, \cdots)$ are also defined in a similar way.

Lemma 3. *For any disjoint sets* $A_1, \cdots, A_m, \{i_0, i_1\} \subset \mathbb{N}$,
i.

$$\mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0\}) > 0$$

ii.

$$\mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0, i_1\}) > 0$$

Proof. *i.*

$$\begin{aligned} \mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0\}) &= \prod_{k=1}^{m} L^*_k(A_k) \left(1 - \sum_{k'=1}^{m} \frac{L^*_{k'}(A_{k'} \cup \{i_0\})}{L^*_{k'}(A_{k'})}\right) \\ &= \prod_{k=1}^{m} L^*_k(A_k) \left(1 - \sum_{k'=1}^{m} \frac{L^*_{k'}(\{i_{1,k'}, i_{2,k'}, i_0\})}{L^*_{k'}(\{i_{1,k'}, i_{2,k'}\})}\right) \end{aligned}$$

where $i_{1,k'}, i_{2,k'} \in A_{k'}$ are two closest elements to i_0 among $A_{k'}$ such that if $\min A_{k'} < i_0 < \max A_{k'}$, then $i_{1,k'} < i_0 < i_{2,k'}$, if $i_0 > \max A_{k'}$, then $i_{1,k'} < i_{2,k'} < i_0$, if $i_0 < \min A_{k'}$, then $i_0 < i_{1,k'} < i_{2,k'}$, and if $A_{k'} = \emptyset$, then $i_{1,k'} = i_{2,k'} = \emptyset$. Therefore,

$$\frac{L^*_{k'}(\{i_{1,k'}, i_{2,k'}, i_0\})}{L^*_{k'}(\{i_{1,k'}, i_{2,k'}\})} = \begin{cases} \dfrac{(p_{k'} + c_{k'}|i_{1,k'} - i_0|^{2H_{k'}-2})(p_{k'} + c_{k'}|i_0 - i_{2,k'}|^{2H_{k'}-2})}{p_{k'} + c_{k'}|i_{1,k'} - i_{2,k'}|^{2H_{k'}-2}} & \text{if } \min A_{k'} < i_0 < \max A_{k'}, \\ p_{k'} + c_{k'}|\max A_{k'} - i_0|^{2H_{k'}-2} & \text{if } i_0 > \max A_{k'}, \\ p_{k'} + c_{k'}|\min A_{k'} - i_0|^{2H_{k'}-2} & \text{if } i_0 < \min A_{k'}, \\ p_{k'} & \text{if } A_{k'} = \emptyset. \end{cases}$$

By (11), $\sum_{k'=1}^m \frac{L^*_{k'}(\{i_{1,k'}, i_{2,k'}, i_0\})}{L^*_{k'}(\{i_{1,k'}, i_{2,k'}\})} < 1$, and the result is derived.

ii. Since,

$$\mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0, i_1\}) = \mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0\}) - \sum_{k=1}^m \mathbf{D}^*(\cdots, A_k \cup \{i_1\}, \cdots; \{i_0\}),$$

it is sufficient if we show:

$$\frac{\mathbf{L}^*(A_1, A_2, \cdots, A_m) - \sum_{k=1}^m \mathbf{L}^*(\cdots, A_k \cup \{i_0\}, \cdots)}{\sum_{k'=1}^m \mathbf{L}^*(\cdots, A_{k'} \cup \{i_1\}, \cdots) - \sum_{k,k'=1}^m \mathbf{L}^*(\cdots, A_k \cup \{i_0\}, A_{k'} \cup \{i_1\}, \cdots)} > 1.$$

Note that:

$$\frac{\mathbf{L}^*(A_1, A_2, \cdots, A_m)}{\sum_{k'=1}^m \mathbf{L}^*(\cdots, A_{k'} \cup \{i_1\}, \cdots)} = \frac{1}{\sum_{k'=1}^m \frac{L^*_{k'}(\{i_{1,k'}, i_{2,k'}, i_0\})}{L^*_{k'}(\{i_{1,k'}, i_{2,k'}\})}},$$

which is non-increasing as set A_k increases for $k = 1, \cdots, m$. That is,

$$\frac{\mathbf{L}^*(A_1, A_2, \cdots, A_m)}{\sum_{k'=1}^m \mathbf{L}^*(\cdots, A_{k'} \cup \{i_1\}, \cdots)} \leq \frac{\mathbf{L}^*(A'_1, A'_2, \cdots, A'_m)}{\sum_{k'=1}^m \mathbf{L}^*(\cdots, A'_{k'} \cup \{i_1\}, \cdots)}$$

for any sets $A_k \subseteq A'_k, k = 1, 2, \cdots, m$. Therefore,

$$\frac{\mathbf{L}^*(A_1, A_2, \cdots, A_m)}{\sum_{k'=1}^m \mathbf{L}^*(\cdots, A_{k'} \cup \{i_1\}, \cdots)} > \frac{\sum_{k=1}^m \mathbf{L}^*(\cdots, A_k \cup \{i_0\}, \cdots)}{\sum_{k,k'=1}^m \mathbf{L}^*(\cdots, A_k \cup \{i_0\}, A_{k'} \cup \{i_1\}, \cdots)}$$

by *iii* of Lemma 2. By *i* of Lemma 2 combined with the fact that:

$$\frac{1}{\sum_{k'=1}^m \frac{L^*_{k'}(\{i_{1,k'}, i_{2,k'}, i_0\})}{L^*_{k'}(\{i_{1,k'}, i_{2,k'}\})}} > 1$$

from (11), the result is derived. □

Note that for any disjoint sets $A_1, A_2, \cdots, A_m, \{i_0, i_1, \cdots, i_n\}$

$$\begin{aligned} \mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0, i_1, \cdots, i_n\}) = {} & \mathbf{D}^*(A_1, A_2, \cdots, A_m; \{i_0, i_1, \cdots, i_{n-1}\}) \\ & - \mathbf{D}^*(A_1 \cup \{i_n\}, A_2, \cdots, A_m; \{i_0, i_1, \cdots, i_{n-1}\}) \\ & - \mathbf{D}^*(A_1, A_2 \cup \{i_n\}, \cdots, A_m; \{i_0, i_1, \cdots, i_{n-1}\}) \\ & \cdots \\ & - \mathbf{D}^*(A_1, A_2, \cdots, A_m \cup \{i_n\}; \{i_0, i_1, \cdots, i_{n-1}\}). \end{aligned}$$

Let us denote:

$$\sum_{k=1}^{m} \mathbf{D}^*(A_1, \cdots, A_{k-1}, A_k \cup \{i_n\}, A_{k+1} \cdots, A_m; \{i_0, i_1, \cdots, i_{n-1}\})$$

by:

$$\sum_{k=1}^{m} \mathbf{D}^*(\cdots, A_k \cup \{i_n\}, \cdots; \{i_0, i_1, \cdots, i_{n-1}\}).$$

Proof of Proposition 1. We will show by mathematical induction that $\{X_{i_1}, \cdots, X_{i_n}\}$ is a random vector with probability (4)–(6) for any n and any $\{i_1, i_2, \cdots, i_n\} \subset \mathbb{N}$. For $n = 1$, it is trivial. For $n = 2$, it is proved by Lemma 3. Let us assume that $\{X_{i_1}, \cdots, X_{i_{n'-1}}\}$ is a random vector with probability (4)–(6) for any $\{i_1, i_2, \cdots, i_{n'-1}\} \subset \mathbb{N}$. We will prove that $\{X_{i_1}, \cdots, X_{i_{n'}}\}$ is a random vector for any $\{i_1, i_2, \cdots, i_{n'}\} \subset \mathbb{N}$.

Without loss of generality, fix a set $\{i_1, i_2, \cdots, i_{n'}\} \subset \mathbb{N}$. To prove that $\{X_{i_1}, \cdots, X_{i_{n'}}\}$ is a random vector with probability (4)–(6), we need to show that $\mathbf{D}^*(A_1, \cdots, A_m; A_0) > 0$ for any pairwise disjoint sets, $A_0, \cdots, A_m$, such that $\cup_{k=0}^{m} A_k = \{i_1, \cdots, i_{n'}\}$. If $|A_0| = 0$ or 1, then the result follows from the definition of $\mathbf{D}^*$ and Lemma 3, respectively. Therefore, we assume that $|A_0| \geq 2$, $A_0 = \{i'_0, i'_1, \cdots, i'_{n_0}\}$, and $\max A_0 = i'_{n_0}$. Let $A'_0 = A_0 / \{i'_{n_0}\}$. We will first show that for any such sets,

$$\frac{\mathbf{D}^*(A_1, \cdots, A_m; A'_0)}{\sum_{\ell=1}^{m} \mathbf{D}^*(\cdots, A_\ell \cup \{i'_{n_0}\}, \cdots; A'_0)} > 1. \tag{13}$$

(13) is equivalent to $\mathbf{D}^*(A_1, \cdots, A_m; A_0) > 0$.

For fixed $\ell \in \{1, 2, \cdots, m\}$, define the following vectors of length $m - 1$,

$$\begin{aligned}
\mathbf{H}^\ell &= (H_1, \cdots, H_{\ell-1}, H_{\ell+1}, \cdots, H_m), \\
\mathbf{p}^\ell &= (p_1, \cdots, p_{\ell-1}, p_{\ell+1}, \cdots, p_m), \\
\mathbf{c}^\ell &= (c_1, \cdots, c_{\ell-1}, c_{\ell+1}, \cdots, c_m).
\end{aligned}$$

We also define:

$$\mathbf{D}^*_{(-\ell)}(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0) := D^*_{\mathbf{H}^\ell, \mathbf{p}^\ell, \mathbf{c}^\ell}(A_1, \cdots, A_{\ell-1}, A_{\ell+1}, \cdots, A_m; A_0).$$

Since $\{X_i; i \in \cup_{k=1}^{m} A_k \cup A'_0\}$ is a random vector with (4)–(6), $\mathbf{D}^*(\cdots, A_\ell, \cdots; A'_0) > 0$, and it can be written as:

$$\mathbf{D}^*(\cdots, A_\ell,, \cdots; A'_0) = P\Big(\cap_{i \in A'_0} \{X_i = 0\} \cap_{\substack{i \in A_k \\ k=1,\cdots,m \\ k \neq \ell}} \{X_i = k\} \cap_{i \in A_\ell} \{X_i = \ell\} \Big) \tag{14}$$

$$= P\Big(\cap_{i\in A_0'} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell} \{X_i = \ell\} \Big)$$

$$- P\Big(\cap_{i\in A_0'/\{i_0'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_0'\}} \{X_i = \ell\} \Big)$$

$$- P\Big(\cap_{i\in A_0'/\{i_0',i_1'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_1'\}} \{X_i = \ell\} \cap \{X_{i_0'} = 0\} \Big)$$

$$- P\Big(\cap_{i\in A_0'/\{i_0',i_1',i_2'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_2'\}} \{X_i = \ell\} \cap_{i\in\{i_0',i_1'\}} \{X_i = 0\} \Big)$$

$$\vdots$$

$$- P\Big(\cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_{n_0-1}'\}} \{X_i = \ell\} \cap_{i\in A_0'/\{i_{n_0-1}'\}} \{X_i = 0\} \Big).$$

Note that:

$$\begin{aligned} & P\Big(\cap_{i\in A_0'} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell} \{X_i = \ell\} \Big) \\ & = P\Big(\cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell} \{X_i = \ell\} \Big) \\ & - P\Big(\cap_{i\in A_0'} \{X_i \in \{1,\cdots,\ell-1,\ell+1,\cdots,m\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell} \{X_i = \ell\} \Big) \\ & = L_\ell^*(A_\ell)\mathbf{D}_{(-\ell)}^*(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0'), \end{aligned} \tag{15}$$

and:

$$\begin{aligned} & P\Big(\cap_{i\in\{i_{j+1}',\cdots,i_{n_0-1}'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \\ & \qquad \cap_{i\in A_\ell\cup\{i_j'\}} \{X_i = \ell\} \cap_{i\in\{i_0',\cdots,i_{j-1}'\}} \{X_i = 0\} \Big) \\ & = P\Big(\cap_{i\in A_0'/\{i_j'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_j'\}} \{X_i = \ell\} \Big) \\ & - \sum_{i^*\in A_0', i^*<i_j'} P\Big(\cap_{i\in A_0'/\{i_j',i^*\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_j',i^*\}} \{X_i = \ell\} \Big) \\ & + \sum_{\substack{i^*,i^{**}\in A_0', \\ i^*<i^{**}<i_j'}} P\Big(\cap_{i\in A_0'/\{i_j',i^*,i^{**}\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_j',i^*,i^{**}\}} \{X_i = \ell\} \Big) \\ & \vdots \\ & (-1)^j P\Big(\cap_{i\in A_0'/\{i_j',i_0',i_1',\cdots,i_{j-1}'\}} \{X_i \in \{0,\ell\}\} \cap_{\substack{i\in A_k \\ k=1,\cdots,m \\ k\neq\ell}} \{X_i = k\} \cap_{i\in A_\ell\cup\{i_j',i_0',i_1',\cdots,i_{j-1}'\}} \{X_i = \ell\} \Big) \\ & = \sum_{\substack{C\cap D=\emptyset \\ C=\emptyset \text{ or } \max C<i_j' \\ C\cup D=A_0'/\{i_j'\}}} (-1)^{|C|} L_\ell^*(A_\ell \cup \{i_j'\} \cup C)\mathbf{D}_{(-\ell)}^*(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; D) \end{aligned} \tag{16}$$

where $|\emptyset| = 0$. Therefore, by (14)–(16),

$$\mathbf{D}^*(\cdots,A_\ell,\cdots;A_0') = L_\ell^*(A_\ell)\mathbf{D}^*_{(-\ell)}(\cdots,A_{\ell-1},A_{\ell+1},\cdots;A_0') \tag{17}$$
$$+\sum_{j=0}^{n_0-1}\sum_{\substack{C\cap D=\emptyset\\ C=\emptyset \text{ or } \max C<i_j'\\ C\cup D=A_0'/\{i_j'\}}}(-1)^{|C|+1}L_\ell^*(A_\ell\cup\{i_j'\}\cup C)\mathbf{D}^*_{(-\ell)}(\cdots,A_{\ell-1},A_{\ell+1},\cdots;D).$$

(17) can also be derived by the definition of $L_\ell^*, \mathbf{D}^*$, without using probability for $\{X_i; i\in \cup_{k=1}^m A_k\cup A_0'\}$. In the same way, using the definition of $L_\ell^*, \mathbf{D}^*$,

$$\mathbf{D}^*(\cdots,A_\ell\cup\{i_{n_0}'\},\cdots;A_0') = L_\ell^*(A_\ell\cup\{i_{n_0}'\})\mathbf{D}^*_{(-\ell)}(\cdots,A_{\ell-1},A_{\ell+1},\cdots;A_0') \tag{18}$$
$$+\sum_{j=0}^{n_0-1}\sum_{\substack{C\cap D=\emptyset\\ C=\emptyset \text{ or } \max C<i_j'\\ C\cup D=A_0'/\{i_j'\}}}(-1)^{|C|+1}L_\ell^*(A_\ell\cup\{i_{n_0}', i_j'\}\cup C)\mathbf{D}^*_{(-\ell)}(\cdots,A_{\ell-1},A_{\ell+1},\cdots;D).$$

Note that, for $j=0,1,\cdots,n_0-1$,

$$g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell\cup\{i_{n_0}'\},\cdots,A_m;A_0';i_j') :=$$
$$\sum_{\substack{C\cap D=\emptyset\\ C=\emptyset \text{ or } \max C<i_j'\\ C\cup D=A_0'/\{i_j'\}}}(-1)^{|C|+1}L_\ell^*(A_\ell\cup\{i_{n_0}', i_j'\}\cup C)\mathbf{D}^*_{(-\ell)}(\cdots,A_{\ell-1},A_{\ell+1},\cdots;D)<0,$$

since we have:

$$g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell,\cdots,A_m;A_0';i_j') =$$
$$-P\Big(\cap_{i\in\{i_{j+1}',\cdots,i_{n_0-1}'\}}\{X_i\in\{0,\ell\}\}\cap_{\substack{i\in A_k\\ k=1,\cdots,m\\ k\neq\ell}}\{X_i=k\}\cap_{i\in A_\ell\cup\{i_j'\}}\{X_i=\ell\}$$
$$\cap_{i\in\{i_0',\cdots,i_{j-1}'\}}\{X_i=0\}\Big)<0$$

by (16), and:

$$f_{H_\ell,p_\ell,c_\ell}(A_\ell;i_j';i_{n_0}') := \frac{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell,\cdots,A_m;A_0';i_j')}{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell\cup\{i_{n_0}'\},\cdots,A_m;A_0';i_j')}>1. \tag{19}$$

The last inequality is due to the fact that:

$$\frac{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell,\cdots,A_m;A_0';i_j')}{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1,\cdots,A_\ell\cup\{i_{n_0}'\},\cdots,A_m;A_0';i_j')}$$
$$=\frac{\sum_{j=0}^{n_0-1}\sum_{\substack{C\subseteq A_0'/\{i_j'\}\\ C=\emptyset \text{ or } \max C<i_j'}}(-1)^{|C|+1}L_\ell^*(A_\ell\cup\{i_j'\}\cup C)}{\sum_{j=0}^{n_0-1}\sum_{\substack{C\subseteq A_0'/\{i_j'\}\\ C=\emptyset \text{ or } \max C<i_j'}}(-1)^{|C|+1}L_\ell^*(A_\ell\cup\{i_{n_0}', i_j'\}\cup C)},$$

and for any set C such that $\max C < i'_j$ or $C = \emptyset$,

$$\frac{L^*_\ell(A_\ell \cup \{i'_j\} \cup C)}{L^*_\ell(A_\ell \cup \{i'_{n_0}, i'_j\} \cup C)} = \frac{L^*_\ell(A_\ell \cup \{i'_j\})}{L^*_\ell(A_\ell \cup \{i'_{n_0}, i'_j\})} > 1$$

by (11). More specifically,

$$f_{H_\ell, p_\ell, c_\ell}(A_\ell; i'_j; i'_{n_0}) = \frac{L^*_\ell(A_\ell \cup \{i'_j\} \cup C)}{L^*_\ell(A_\ell \cup \{i'_{n_0}, i'_j\} \cup C)} = \frac{L^*_\ell(i_{\ell,j,1}, i_{\ell,j,2})}{L^*_\ell(i_{\ell,j,1}, i_{\ell,j,2}, i'_{n_0})} \tag{20}$$

where $i_{\ell,j,1}, i_{\ell,j,2}$ are the two closest elements to i'_{n_0} among $A_\ell \cup \{i'_j\}$. That is, $i_{\ell,j,1}, i_{\ell,j,2} \in A_\ell \cup \{i'_j\}$ are two closest elements to i'_{n_0} such that if $\min A_\ell \cup \{i'_j\} < i'_{n_0} < \max A_\ell$, then $i_{\ell,j,1} < i'_{n_0} < i_{\ell,j,2}$, and if $i'_{n_0} > \max A_\ell \cup \{i'_j\}$, then $i_{\ell,j,1} < i_{\ell,j,2} < i'_{n_0}$.

$$\begin{aligned}
&\frac{L^*_\ell(\{i_{\ell,j,1}, i_{\ell,j,2}\})}{L^*_\ell(\{i_{\ell,j,1}, i_{\ell,j,2}, i_{n'}\})} \\
&= \begin{cases} \dfrac{p_\ell + c_\ell |i_{\ell,j,1} - i_{\ell,j,2}|^{2H_\ell - 2}}{(p_\ell + c_\ell |i_{\ell,j,1} - i_{n'}|^{2H_\ell - 2})(p_\ell + c_\ell |i_{n'} - i_{\ell,j,2}|^{2H_\ell - 2})} & \text{if } \min A_\ell \cup \{i'_j\} < i_{n'} < \max A_\ell, \\ \dfrac{1}{p_\ell + c_\ell |i_{\ell,j,2} - i_{n'}|^{2H_\ell - 2}} & \text{if } i_{n'} > \max A_\ell \cup \{i'_j\}, \end{cases}
\end{aligned}$$

which is non-increasing as j increases since $i'_j < i'_{n_0}$. Therefore, $f_{H_\ell, p_\ell, c_\ell}(A_\ell; i'_j; i'_{n_0})$ is non-increasing as j increases. Also, for fixed j, C such that $\max C < i'_j$ or $C = \emptyset$,

$$\frac{L^*_\ell(A_\ell \cup \{i'_{n_0}, i'_j\} \cup C)}{L^*_\ell(A_\ell \cup \{i'_j\} \cup C)} \geq \frac{L^*_\ell(A_\ell \cup \{i'_{n_0}\})}{L^*_\ell(A_\ell)} \tag{21}$$

by the fact that $\frac{L^*_\ell(A \cup \{i\})}{L^*_\ell(A)}$ is non-decreasing as the set A increases.

Combining the above facts with (17) and (18), and by i of Lemma 2,

$$\frac{L^*_\ell(A_\ell)}{L^*_\ell(A_\ell \cup \{i'_{n_0}\})} \leq \frac{\mathbf{D}^*(\cdots, A_\ell, \cdots; A'_0)}{\mathbf{D}^*(\cdots, A_\ell \cup \{i'_{n_0}\}, \cdots; A'_0)}.$$

Therefore,

$$\frac{\mathbf{D}^*(A_1, \cdots, A_m; A'_0)}{\sum_{\ell=1}^m \mathbf{D}^*(\cdots, A_\ell \cup \{i'_{n_0}\}, \cdots; A'_0)} \geq \frac{1}{\sum_{\ell=1}^m \frac{L^*_\ell(A_\ell \cup \{i'_{n_0}\})}{L^*_\ell(A_\ell)}} > 1,$$

which proves (13) and,

$$\mathbf{D}^*(A_1, \cdots, A_m; A_0) > 0.$$

□

Proof of Theorem 3. a. Let $A_0 = \{i_0, i_1, \cdots, i_n\}$. Note that:

$$\begin{aligned}
&P(X_{i'_1} = \ell | \cap_{k=0,\cdots,m} \cap_{i \in A_k} \{X_i = k\}) = \frac{\mathbf{D}^*(\cdots, A_\ell \cup \{i'_1\}, \cdots; A_0)}{\mathbf{D}^*(A_1, \cdots, A_m; A_0)} = \\
&\frac{L^*_\ell(A_\ell \cup \{i'_1\}) \mathbf{D}^*_{(-\ell)}(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0) + \sum_{j=0}^n g_{\mathbf{H},\mathbf{p},\mathbf{c}}(\cdots, A_\ell \cup \{i'_1\}, \cdots; A_0; i_j)}{L^*_\ell(A_\ell) \mathbf{D}^*_{(-\ell)}(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0) + \sum_{j=0}^n g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, \cdots, A_m; A_0; i_j)}.
\end{aligned}$$

Since,

$$\frac{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, \cdots, A_\ell \cup \{i'_1\}, \cdots, A_m; A_0; i_j)}{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, \cdots, A_m; A_0; i_j)}$$

is non-decreasing as j increases, and by (19) and (20):

$$\frac{L_\ell^*(A_\ell \cup \{i_1'\})}{L_\ell^*(A_\ell)} \leq \frac{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, \cdots, A_\ell \cup \{i_1'\}, \cdots, A_m; A_0; i_j)}{g_{\mathbf{H},\mathbf{p},\mathbf{c}}(A_1, \cdots, A_\ell, \cdots, A_m; A_0; i_j)},$$

the result follows by *ii* of Lemma 2.
b.

$$\frac{P(X_{i_2'} = \ell | \cap_{k=0,\cdots,m} \cap_{i \in A_k} \{X_i = k\})}{P(X_{i_3'} = \ell | \cap_{k=0,\cdots,m} \cap_{i \in A_k} \{X_i = k\})} = \frac{\mathbf{D}^*(\cdots, A_\ell \cup \{i_2'\}, \cdots; A_0)}{\mathbf{D}^*(\cdots, A_\ell \cup \{i_3'\}, \cdots; A_0)} =$$
$$\frac{L_\ell^*(A_\ell \cup \{i_2'\})\mathbf{D}^*_{(-\ell)}(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0) + \sum_{j=0}^n g_{\mathbf{H},\mathbf{p},\mathbf{c}}(\cdots, A_\ell \cup \{i_2'\}, \cdots, ; A_0; i_j)}{L_\ell^*(A_\ell \cup \{i_3'\})\mathbf{D}^*_{(-\ell)}(\cdots, A_{\ell-1}, A_{\ell+1}, \cdots; A_0) + \sum_{j=0}^n g_{\mathbf{H},\mathbf{p},\mathbf{c}}(\cdots, A_\ell \cup \{i_3'\}, \cdots; A_0; i_j)}.$$

For fixed j, C such that $\max C < i_j$,

$$\frac{L_\ell^*(A_\ell \cup \{i_2', i_j\} \cup C)}{L_\ell^*(A_\ell \cup \{i_3', i_j\} \cup C)} \leq \frac{L_\ell^*(A_\ell \cup \{i_2'\})}{L_\ell^*(A_\ell \cup \{i_3'\})},$$

and,

$$\frac{L_\ell^*(A_\ell \cup \{i_2', i_j\} \cup C)}{L_\ell^*(A_\ell \cup \{i_3', i_j\} \cup C)}$$

is non-increasing as j increases. Therefore, the result follows by *i* of Lemma 2. □

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Hurst, H. Long-term storage capacity of reservoirs. *Civ. Eng. Trans.* **1951**, *116*, 770–808. [CrossRef]
2. Hurst, H. Methods of using long-term storage in reservoirs. *Proc. Inst. Civ. Eng.* **1956**, *5*, 519–543. [CrossRef]
3. Benson, D.A.; Meerschaert, M.M.; Baeumer, B.; Scheffler, H.-P. Aquifer operator-scaling and the effect on solute mixing and dispersion. *Water Resour. Res.* **2006**, *42*, W01415. [CrossRef]
4. Delgado, R. A reflected fBm limit for fluid models with ON/OFF sources under heavy traffic. *Stoch. Processes Their Appl.* **2007**, *117*, 188–201. [CrossRef]
5. Majewski, K. Fractional Brownian heavy traffic approximations of multiclass feedforward queueing networks. *Queueing Syst.* **2005**, *50*, 199–230. [CrossRef]
6. Samorodnitsky, G. *Stochastic Processes and Long Range Dependence*; Springer: Berlin/Heidelberg, Germany, 2016.
7. Daley, J.; Vesilo, R. Long range dependence of point processes, with queueing examples. *Stoch. Processes Their Appl.* **1997**, *70*, 265–282. [CrossRef]
8. Daley, J.; Vesilo, R. Long range dependence of inputs and outputs of classical queues. *Fields Inst. Commun.* **2000**, *28*, 179–186.
9. Mandelbrot, B.; Van Ness, J. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **1968**, *10*, 422–437. [CrossRef]
10. Hosking, J.R.M. Fractional differencing. *Biometrika* **1981**, *68*, 165–176. [CrossRef]
11. Hosking, J.R.M. Modeling persistence in hydrological time series using fractional differencing. *Water Resour. Res.* **1984**, *20*, 1898–1908. [CrossRef]
12. Carpio, K.J.E. Long-Range Dependence of Markov Chains. Ph.D. Thesis, The Australian National University, Canberra, Australia, 2006.
13. Dean, C.B.; Lundy, E.R. Overdispersion. Wiley StatsRef: Statistics Reference Online. 2014. Available online: https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06788.pub2 (accessed on 9 October 2022).
14. Poortema, K. On modelling overdispersion of counts. *Stat. Neerl.* **1999**, *53*, 5–20. [CrossRef]
15. Afroz, F. Estimating Overdispersion in Sparse Multinomial Data. Ph.D. Thesis, The University of Otago, Dunedin, New Zealand, 2018.
16. Afroz, F.; Shabuz, Z.R. Comparison Between Two Multinomial Overdispersion Models Through Simulation. *Dhaka Univ. J. Sci.* **2020**, *68*, 45–48. [CrossRef]
17. Landsman, V.; Landsman, D.; Bang, H. Overdispersion models for correlated multinomial data: Applications to blinding assessment. *Stat. Med.* **2019**, *38*, 4963–4976. [CrossRef] [PubMed]

18. Mosimann, J.E. On the Compound Multinomial Distribution, the Multivariate β- Distribution, and Correlations Among Proportions. *Biometrika* **1962**, *49*, 65–82.
19. Lee, J. Generalized Bernoulli process and fractional binomial distribution. *Depend. Model.* **2021**, *9*, 1–12. [CrossRef]
20. Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; John Wiley: New York, NY, USA, 1968; Volume 1.
21. Lee, J. Generalized Bernoulli process and fractional binomial distribution II. *arXiv* **2022**, arXiv:2209.01516
22. Carpio, K.J.E.; Daley, D.J. Long-Range Dependence of Markov Chains in Discrete Time on Countable State Space. *J. Appl. Probab.* **2007**, *44*, 1047–1055. [CrossRef]