Number of Distinct Values in a Large Sample with Dependent Observations Generated by Fractional Gaussian Noise from an Infinite Discrete Distribution

N. S. Arkashov [*, a]

[a]Sobolev Institute of Mathematics, Acad. Koptyug ave., 4, 630090, Novosibirsk, Russia.

Abstract

The dynamics of growth of the number of distinct values in consecutive samples obtained from a stationary sequence of dependent observations with an infinite discrete distribution is studied. The problem of studying the mentioned behavior, where samples are formed from a sequence of i.i.d. random variables, is well-known. In this paper, the expected number of distinct sample values in the independent case is compared with that in the case of dependent observations. A connection is established between the estimation of these expectations and the problem of estimating multivariate normal distributions.

Keywords: urn scheme, fractional noise, transform of Gaussian sequence, long-range dependence.

1 Introduction

Consider the following urn scheme. Let n balls be thrown into an infinite array of cells, and the probability of each ball hitting j-th cell is $p_j,\ j = 1, 2, \dots$ (assume that $p_j > 0$ for all j). By X_k we denote the number of the cell

[*]Corresponding author.
E-mail: nicky1978@mail.ru (N. S. Arkashov).

into which the k-ball hits ($k = 1, \dots, n$), as a result we obtain a sample of identically distributed random variables: $(X_1, \dots, X_n)$, where X_1 has a discrete distribution with atoms at points $1, 2, \dots$ and probabilities $p_1, p_2, \dots$. Let F denote the cdf corresponding to the mentioned discrete distribution. Denote by T_n the number of distinct values in $(X_1, \dots, X_n)$. Note that when $(X_1, \dots, X_n)$ are independent (which corresponds to n balls being thrown independently of each other), as $n \to +\infty$, the law of large numbers and the central limit theorem hold for T_n, as established in [1] and [2], respectively.

In this paper, a stationary (in the strict sense) sequence $\{X_k,\ k = 1, 2, \dots\}$ such that X_1 has a distribution specified by F is defined constructively, and the behavior of T_n is investigated for this sequence. In particular, it is proved that $T_n \to +\infty$ almost surely (a.s.).

Additionally, we compare $\mathbf{E}T_n^{(1/2)}$ and $\mathbf{E}T_n^{(H)}$ ($H \neq 1/2$), where the superscripts denote the independent and dependent cases for the sequence $\{X_k\}$, respectively (a detailed definition is provided below).

2 Main results

2.1 Preliminaries

We denote by F^{-1} the quantile transformation of the function F, defined as $F^{-1}(t) := \inf\{x : F(x) \geq t\}$. Let $\{z_k\}$ be a standard fractional noise with parameter $H \in (0, 1)$, i.e., a centered Gaussian sequence with covariance function

$$\rho(j) := \frac{1}{2}(|j+1|^{2H} + |j-1|^{2H} - 2|j|^{2H}),\ j \geq 0. \tag{1}$$

If it is important to emphasize that $\{z_k\}$ is the fractional noise with parameter H, we will add the index (H): $\{z_k^{(H)}\}$.

We consider the sequence

$$X_k := F^{-1}(\Phi(z_k)),\ k = 1, 2, \dots, \tag{2}$$

where Φ is the cdf of the standard normal law. Note that $\Phi(z_k) \overset{d}{\sim} U[0, 1],\ k = 1, 2, \dots$, and hence, $X_k,\ k = 1, 2, \dots$ follow the distribution specified by F.

The constructed $\{X_k\}$ is a stationary (in the strict sense) sequence of random variables. In the case $H = 1/2$ this sequence becomes a sequence of independent random variables.

In cases where it is important to emphasize that T_n corresponds to $\{X_k\}$ formed by the fractional noise with parameter H, we will use this notation: $T_n^{(H)}$.

Remark 1. In the case $H > 1/2$, the following holds for the covariance function of the fractional noise: $\rho(k) \sim H(2H-1)k^{2H-2}$, $k \to +\infty$ (e.g., see [3, Proposition 7.2.10]). Suppose that the distribution specified by F possesses a finite second moment. We establish that $\mathbf{Cov}(X_1, X_{k+1})$ and $\rho(k)$ have the same asymptotic order as $k \to +\infty$. By F_1 we denote the cdf of $(X_1 - a)/\sigma$, where $a := \mathbf{E}X_1$, $\sigma^2 := \mathbf{Var}(X_1)$. In accordance with item 3 of Theorem 1 in [4], we derive that $\mathbf{Cov}(X_1, X_{k+1}) \sim \sigma^2\varsigma^2\rho(k)$, where $\varsigma := \int_{-\infty}^{\infty} xF_1^{-1}(\Phi(x))\varphi(x)\,dx$ (here φ is the pdf of the standard normal law). In [4] it is proved that $\varsigma > 0$. This asymptotic behavior of $\mathbf{Cov}(X_1, X_{k+1})$ implies the so-called long-range dependence of $\{X_k\}$ (e.g., see [5, 6]).

2.2 Theorems

Theorem 1. *It holds that $T_n \to +\infty$ as $n \to \infty$ (a.s.).*

Theorem 1, by virtue of Fatou's lemma, implies immediately that $\mathbf{E}T_n \to +\infty$ as $n \to +\infty$.

Next we proceed to compare $\mathbf{E}T_n^{(H)}$ $(H > 1/2)$ and $\mathbf{E}T_n^{(1/2)}$, for this purpose we will use the following lemma.

Lemma 1. *It holds that*

$$\mathbf{E}T_n = \sum_{j=1}^{+\infty}(1 - \mathbf{P}(z_1 \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)], \dots, z_n \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)])),$$

where $P_0 := 0$, $P_j := \sum_{i=1}^{j} p_i$, $j \geq 1$.

Thus, it follows from Lemma 1 that the problem of comparing $\mathbf{E}T_n^{(H)}$ $(H \neq 1/2)$ and $\mathbf{E}T_n^{(1/2)}$ reduces to the problem of comparing $\mathbf{P}(z_1^{(H)} \notin (c,d], \dots, z_n^{(H)} \notin (c,d])$ and $\mathbf{P}(z_1^{(1/2)} \notin (c,d], \dots, z_n^{(1/2)} \notin (c,d])$, where $c < d$, $c \in \mathbb{R} \cup \{-\infty\}$ and $d \in \mathbb{R}$.

Observe that for $H > 1/2$, the inequality $\mathbf{Cov}(z_i^{(H)}, z_j^{(H)}) \geq \mathbf{Cov}(z_i^{(1/2)}, z_j^{(1/2)})$ holds for each pair i, j. From this, in the case $c = -\infty$, we have for any $d \in \mathbb{R}$:

$$\mathbf{P}(z_1^{(H)} \notin (c,d], \dots, z_n^{(H)} \notin (c,d]) \geq \mathbf{P}(z_1^{(1/2)} \notin (c,d], \dots, z_n^{(1/2)} \notin (c,d]). \quad (3)$$

This fact immediately follows from [7, Lemma 1] (see also [8, Lemma 4.2.3]). Note also the monograph [9] devoted to multivariate normal distributions, in particular, this monograph deals with inequalities of the form (3). A similar claim for finite c, d is proved in the present paper only in the case $n = 2$ (see Proposition 1 and Corollary 1).

Theorem 2. *Let $H > 1/2$; then, $\mathbf{E}T_2^{(H)} < \mathbf{E}T_2^{(1/2)}$.*

Note that in the proof of this theorem, the condition $H > 1/2$ plays a crucial role, as it ensures the positivity of the covariance function $\rho(\cdot)$ (see the proof of Corollary 1).

2.3 Statistical illustration and some assumptions

We will simulate $n = 5000$ independent samples of the fractional noise with parameter $H = 0.9$ and size n: $Z_k := (z_{k,1}, \dots, z_{k,n})$, $k = 1, \dots, n$. Let F be specified by a discrete distribution with atoms at points $i = 1, 2, \dots$ and probabilities $p_i := b/i^4$, where b is the corresponding normalization constant. Based on Z_k, $k = 1, \dots, n$, we construct $X_{k,i} := F^{-1}(\Phi(z_{k,i}))$, $i = 1, \dots, n$, $k = 1, \dots, n$ (see (2)).

Denote by $T_{k,j}$ ($j = 1, \dots, n$, $k = 1, \dots, n$) the number of distinct values of $(X_{k,1}, \dots, X_{k,j})$. Set

$$\overline{T}_{n,j} := \frac{1}{n} \sum_{k=1}^{n} T_{k,j}, \ \ j = 1, 2, \dots . \tag{4}$$

Note that $\overline{T}_{n,j}$ is a consistent estimator for $\mathbf{E}T_j^{(0.9)}$.

In the case $H = 1/2$, the value of $\mathbf{E}T_j^{(1/2)}$ is of the form (see [1])

$$\mathbf{E}T_j^{(1/2)} = \sum_{k=1}^{+\infty} (1 - (1 - p_k)^j).$$

Consider the plots of $y_1(j) := \mathbf{E}T_j^{(1/2)}$ and $y_2(j) := \overline{T}_{n,j}$, $j = 1, \dots, n$ (see Figure 1).

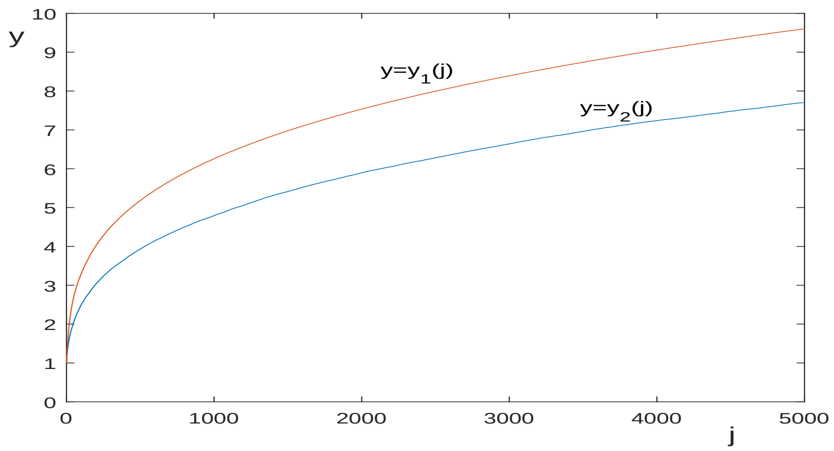

Fig. 1. $y_1(j) = \mathbf{E}T_j^{(1/2)}$, $y_2(j) = \overline{T}_{n,j}$, $p_i = b/i^4$, $i = 1, 2, \dots$.

Let us give an estimate of the proximity of y_1 to y_2 relative to y_1 (corresponding to the case $H = 1/2$). We have $\Delta_1 := \max_j (y_1(j) - y_2(j))/y_1(j) \approx 0.25$.

Consider the situation where $\{p_i\}$ decays to 0 at a slower rate than in the above case. Let $p_i := b/(i+q)^s$, $i = 1, 2, \dots$, where $q = 2.7$, $s = 1.5$, b is the normalization constant.

Next, as before, we form $\{X_{k,i}\}$ on the basis of the fractional noise with the parameter $H = 0.9$. The corresponding plots of $y_1(j) := \mathbf{E}T_j^{(1/2)}$ and $y_2(j) := \overline{T}_{n,j}$, $j = 1, \dots, n$ are shown in Figure 2.

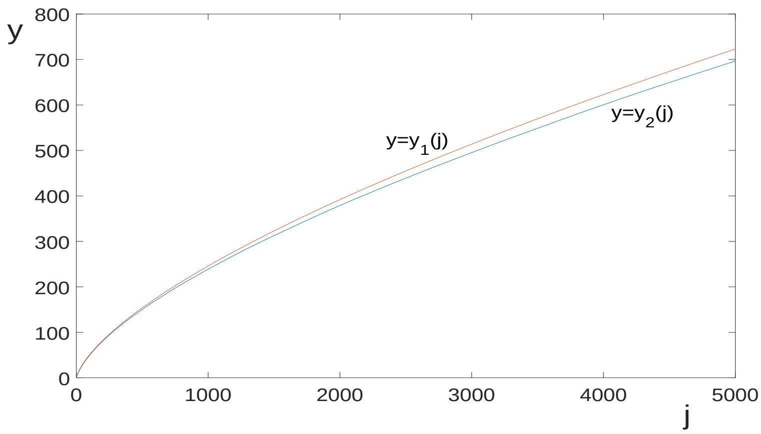

Fig. 2. $y_1(j) = \mathbf{E}T_j^{(1/2)}$, $y_2(j) = \overline{T}_{n,j}$, $p_i = b/(i+2.7)^{1.5}$, $i = 1, 2, \dots$.

The estimate of the proximity of y_1 to y_2 (relative to y_1) in this case is $\Delta_2 := \max_j (y_1(j) - y_2(j))/y_1(j) \approx 0.07$. As a result, we obtain that Δ_2 is significantly smaller than Δ_1. In connection with this, it is noteworthy that the slower decay of $\{p_i\}$ to 0 (compared to the first case) has a significant impact on the growth dynamics of $\mathbf{E}T_j^{(0.9)}$, which is substantially greater than the influence of the dependence structure of $\{X_k\}$.

According to Figure 1 and Figure 2, we may assume that $\mathbf{E}T_j^{(1/2)} > \mathbf{E}T_j^{(0.9)}$ for all $j = 1, \dots, n$ (see also the remark to the proof of Theorem 2 in Subsection 3.2).

3 Proofs

3.1 Proof of Theorem 1

Below, we use the following lemma (e.g., see [8]).

Lemma 2. *Let $\{\xi_n\}$ be a stationary sequence of standard normal random variables with covariances $\{r_n\}$ satisfying the condition $r_n \ln n \to 0$. Then, for each $x \in \mathbb{R}$, it holds that*

$$\lim_{n\to+\infty} \mathbf{P}(a_n(\max_{1\le i\le n} \xi_i - b_n) \le x) = \exp(-e^{-x}),$$

where $a_n = 2(\ln n)^{1/2}$, $b_n = a_n - (2a_n)^{-1}(\ln\ln n + \ln 4\pi)$.

For almost every ω, the sequence $\{T_n(\omega)\}$ is increasing; therefore, for almost every ω, there exists $\lim T_n(\omega)$ (finite or infinite). Consider $B := \{\omega : \lim T_n(\omega) < +\infty\}$. It holds that

$$B \subseteq \bigcup_{j=1}^{+\infty} \{X_1 \le j, X_2 \le j, \dots\}.$$

Recall the notation: $P_0 = 0$, $P_j = \sum_{i=1}^{j} p_i$, $j \ge 1$. Note that $\{X_k \le j\}$ coincides with $\{z_k \in (-\infty, \Phi^{-1}(P_j)]\}$.

Let us estimate $\mathbf{P}(X_1 \le j, X_2 \le j, \dots)$. The following relations are satisfied:

$$\begin{aligned}&\mathbf{P}(X_1 \le j, X_2 \le j, \dots)\\ &= \mathbf{P}(z_1 \le \Phi^{-1}(P_j), z_2 \le \Phi^{-1}(P_j), \dots) = \lim_{n\to+\infty} \mathbf{P}(\max_{1\le i\le n} z_i \le \Phi^{-1}(P_j)).\end{aligned} \tag{5}$$

In the case $H = 1/2$, it immediately follows from (5) that $\mathbf{P}(X_1 \le j, X_2 \le j, \dots) = 0$; therefore $\mathbf{P}(B) = 0$ (note that this case is discussed in [1]).

Now let $H \ne 1/2$. Consider the right-hand side of the last equality in (5). Let ε be an arbitrarily small positive number. Choose x such that $\exp(-e^{-x}) \le \varepsilon$. Given that $a_n \to +\infty$ and $b_n \to +\infty$ as $n \to +\infty$ (where a_n, b_n are defined in Lemma 2), we deduce that for all sufficiently large n, the following inequality holds

$$\mathbf{P}(\max_{1\le i\le n} z_i \le \Phi^{-1}(P_j)) \le \mathbf{P}(\max_{1\le i\le n} z_i \le \frac{x}{a_n} + b_n). \tag{6}$$

From (6) and Lemma 2, given that $\rho(j) \sim H(2H-1)j^{2H-2}$, $j \to +\infty$ (e.g., see [3, Proposition 7.2.10]), we obtain the relation

$$\limsup_{n\to+\infty} \mathbf{P}(\max_{1\le i\le n} z_i \le \Phi^{-1}(P_j)) \le \exp(-e^{-x}) \le \varepsilon.$$

Thus, $\lim_{n\to+\infty} \mathbf{P}(\max_{1\le i\le n} z_i \le \Phi^{-1}(P_j)) = 0$, from which we deduce: $\mathbf{P}(X_1 \le j, X_2 \le j, \dots) = 0$ (see (5)). As a result, we conclude that $\mathbf{P}(B) = 0$. The theorem is proved. $\square$

3.2 Proof of Theorem 2

Prior to proceeding, we prove the following statements: Proposition 1, Corollary 1, Lemma 1 and Lemma 3.

Note that the proof of Proposition 1 follows the scheme of the proof of Theorem 4.2.1 of [8]. We also note the work [10], which considers assertions similar to Proposition 1 below.

Proposition 1. *Let $(\xi_1, \dots, \xi_n)$ and $(\eta_1, \dots, \eta_n)$ be Gaussian vectors of standard normal variables with positive definite covariance matrices $\Lambda^1 = (\lambda^1_{ij})$ and $\Lambda^0 = (\lambda^0_{ij})$, respectively. Then for any $c, d \in \mathbb{R}$ such that $c < d$, the following relation holds*:

$$\begin{aligned} &\mathbf{P}(\xi_j \notin (c,d] \text{ for } j = 1,2,\dots,n) - \mathbf{P}(\eta_j \notin (c,d] \text{ for } j = 1,2,\dots,n) \\ &= \sum_{i<j} (\lambda^1_{ij} - \lambda^0_{ij}) \int_0^1 \left(\int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} \Delta_\delta(y_i = c, y_j = d)\, dy' \right) d\delta, \end{aligned} \tag{7}$$

where $\Delta_\delta(y_i = c, y_j = d) := f_\delta(y_i = c, y_j = c) + f_\delta(y_i = d, y_j = d) - 2f_\delta(y_i = c, y_j = d)$. In the previous relation, f_δ denotes the n-dimensional normal density corresponding to $\Lambda^\delta = (\lambda^\delta_{ij})$, where $\Lambda^\delta := \delta\Lambda^1 + (1-\delta)\Lambda^0$, $0 \le \delta \le 1$, and $f_\delta(y_i = c, y_j = d)$ is a function of $n-2$ variables that is obtained from $f_\delta(y_1, \dots, y_n)$, if we set $y_i = c$, $y_j = d$ (thus, the integration in (7) is over all variables $y_1, \dots, y_n$ except y_i, y_j).

Specifically, for $n = 2$, it holds that

$$\begin{aligned} &\mathbf{P}(\xi_j \notin (c,d] \text{ for } j = 1,2) - \mathbf{P}(\eta_j \notin (c,d] \text{ for } j = 1,2) \\ &= (\lambda^1_{12} - \lambda^0_{12}) \int_0^1 (f_\delta(c,c) + f_\delta(d,d) - 2f_\delta(c,d))\, d\delta. \end{aligned} \tag{8}$$

Proof. We have

$$\mathbf{P}(\xi_j \notin (c,d] \text{ for } j = 1,2,\dots,n) = \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} f_1(y_1, \dots, y_n)\, dy$$

and

$$\mathbf{P}(\eta_j \notin (c,d] \text{ for } j = 1,2,\dots,n) = \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} f_0(y_1, \dots, y_n)\, dy.$$

Define $F(\delta)$ as follows:

$$F(\delta) := \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} f_\delta(y_1, \dots, y_n)\, dy.$$

The left part (7) is equal to $F(1) - F(0)$. It is obvious that

$$F(1) - F(0) = \int_0^1 F'(\delta)\, d\delta,$$

where

$$F'(\delta) = \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} \frac{\partial f_\delta(y_1, \dots, y_n)}{\partial \delta}\, dy. \tag{9}$$

The density f_δ depends on δ through λ_{ij}^δ $(i < j)$, noting that $\lambda_{ii}^\delta = 1$ (recall that $\Lambda_\delta = \delta\Lambda^1 + (1-\delta)\Lambda^0$). From (9) we deduce

$$\begin{aligned} F'(\delta) =& \sum_{i<j} \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} \frac{\partial f_\delta}{\partial \lambda_{ij}^\delta} \frac{\lambda_{ij}^\delta}{\partial \delta}\, dy \\ &= \sum_{i<j} (\lambda_{ij}^1 - \lambda_{ij}^0) \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} \frac{\partial f_\delta}{\partial \lambda_{ij}^\delta}\, dy. \end{aligned} \tag{10}$$

The following equality holds (see the proof of Theorem 4.2.1 in [8])

$$\frac{\partial f_\delta}{\partial \lambda_{ij}^\delta} = \frac{\partial^2 f_\delta}{\partial y_i \partial y_j}.$$

Consequently,

$$F'(\delta) = \sum_{i<j} (\lambda_{ij}^1 - \lambda_{ij}^0) \int_{\mathbb{R}\setminus[c,d)} \cdots \int_{\mathbb{R}\setminus[c,d)} \frac{\partial^2 f_\delta}{\partial y_i \partial y_j}\, dy.$$

By integrating with respect to y_i and y_j, and then with respect to δ (from 0 to 1), we obtain (7). $\square$

Corollary 1. *Let (ξ_1, ξ_2) and (η_1, η_2) be two-dimensional Gaussian vectors consisting of standard normal random variables with positive definite covariance matrices $\Lambda^1 = (\lambda_{ij}^1)$ and $\Lambda^0 = (\lambda_{ij}^0)$, respectively. Let, in addition, $\lambda_{12}^1 > \lambda_{12}^0 \geq 0$. Then for any $c, d \in \mathbb{R}$ such that $c < d$:*

$$\mathbf{P}(\xi_1 \notin (c,d], \xi_2 \notin (c,d]) > \mathbf{P}(\eta_1 \notin (c,d], \eta_2 \notin (c,d]). \tag{11}$$

Proof. We will use Proposition 1. Recall that by f_δ, $0 \leq \delta \leq 1$ we denote the 2-dimensional normal density corresponding to $\Lambda^\delta = \delta\Lambda^1 + (1-\delta)\Lambda^0$. We prove that for all $\delta \in [0,1]$

$$f_\delta(c,c) + f_\delta(d,d) - 2f_\delta(c,d) > 0. \tag{12}$$

Jensen's inequality yields

$$
\begin{aligned}
\frac{f_\delta(c,c)+f_\delta(d,d)}{2} &= \frac{\exp(-\frac{2c^2-2\lambda_{12}^\delta c^2}{2(1-(\lambda_{12}^\delta)^2)})}{4\pi\sqrt{1-(\lambda_{12}^\delta)^2}} + \frac{\exp(-\frac{2d^2-2\lambda_{12}^\delta d^2}{2(1-(\lambda_{12}^\delta)^2)})}{4\pi\sqrt{1-(\lambda_{12}^\delta)^2}} \\
> \frac{\exp(-\frac{1}{2(1-(\lambda_{12}^\delta)^2)}\frac{(2c^2-2\lambda_{12}^\delta c^2)+(2d^2-2\lambda_{12}^\delta d^2)}{2})}{2\pi\sqrt{1-(\lambda_{12}^\delta)^2}} &= \frac{\exp(-\frac{(1-\lambda_{12}^\delta)(c^2+d^2)}{2(1-(\lambda_{12}^\delta)^2)})}{2\pi\sqrt{1-(\lambda_{12}^\delta)^2}}.
\end{aligned}
\tag{13}
$$

Next, notice that $(1-\lambda_{12}^\delta)(c^2+d^2) \le c^2+d^2-2\lambda_{12}^\delta cd$ (this follows from the fact that $\lambda_{12}^\delta(c-d)^2 \ge 0$). Therefore,

$$
\frac{1}{2\pi\sqrt{1-(\lambda_{12}^\delta)^2}} \exp\left(-\frac{(1-\lambda_{12}^\delta)(c^2+d^2)}{2(1-(\lambda_{12}^\delta)^2)}\right) \ge f_\delta(c,d).
$$

From the last inequality and (13) follows (12). Applying (12) to (8), we obtain the conclusion of the corollary. □

We are going to use the following result from [7, Lemma 1] (see also [8, Lemma 4.2.3]).

Lemma 3. *Let* $(\xi_1,\ldots,\xi_n)$ *and* $(\eta_1,\ldots,\eta_n)$ *be Gaussian vectors of standard normal variables, and* $\mathbf{Cov}(\xi_i,\xi_j) \le \mathbf{Cov}(\eta_i,\eta_j)$ *for each pair* i, j. *Then for any* $u_1,\ldots,u_n$

$$
\mathbf{P}(\xi_j \le u_j \text{ for } j=1,\ldots,n) \le \mathbf{P}(\eta_j \le u_j \text{ for } j=1,\ldots,n).
$$

Let us prove Lemma 1 formulated in Section 2.2.

Proof of Lemma 1. Define $\{Z_j\}$ by

$$
Z_j = \begin{cases} 1, & \text{if at least one of the random variables } X_1,\ldots,X_n \text{ takes the value } j, \\ 0, & \text{otherwise.} \end{cases}
$$

Clearly, $T_n = \sum_{j=1}^{+\infty} Z_j$, from which we deduce

$$
\mathbf{E}T_n = \sum_{j=1}^{+\infty}(1-\mathbf{P}(X_1 \ne j,\ldots,X_n \ne j)). \tag{14}
$$

It follows from (2) that $\{X_k \ne j\} = \{z_k \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)]\}$. Applying this relation to (14), we obtain the conclusion of the Lemma.

□

Proceed to the proof of Theorem 2. Let Λ^1 be the covariance matrix of the two-dimensional vector corresponding to the fractional noise with parameter $H > 1/2$:

$$\Lambda^1 = \begin{pmatrix} 1 & 2^{2H-1}-1 \\ 2^{2H-1}-1 & 1 \end{pmatrix} \tag{15}$$

and Λ^0 be the covariance matrix of the vector corresponding to the fractional noise with parameter $H = 1/2$:

$$\Lambda^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{16}$$

From Lemma 1 (given that $\Phi^{-1}(0) = -\infty$) it follows that

$$\begin{aligned} &\mathbf{E}T_2^{(H)} = 1 - \mathbf{P}(-z_1^{(H)} < -\Phi^{-1}(p_1), -z_2^{(H)} < -\Phi^{-1}(p_1)) \\ &+ \sum_{j=2}^{+\infty}(1 - \mathbf{P}(z_1^{(H)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)], z_2^{(H)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)])). \end{aligned} \tag{17}$$

From Lemma 3, taking into account (15) and (16), we conclude that

$$\begin{aligned} &\mathbf{P}(-z_1^{(H)} < -\Phi^{-1}(p_1), -z_2^{(H)} < -\Phi^{-1}(p_1)) \\ &\geq \mathbf{P}(-z_1^{(1/2)} < -\Phi^{-1}(p_1), -z_2^{(1/2)} < -\Phi^{-1}(p_1)). \end{aligned} \tag{18}$$

Using Corollary 1 (again, considering (15) and (16)), we get

$$\begin{aligned} &\mathbf{P}(z_1^{(H)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)], z_2^{(H)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)]) \\ &> \mathbf{P}(z_1^{(1/2)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)], z_2^{(1/2)} \notin (\Phi^{-1}(P_{j-1}), \Phi^{-1}(P_j)]). \end{aligned} \tag{19}$$

The assertion of the theorem immediately follows from (18) and (19). □

We note that a generalization of Theorem 2 to compare $\mathbf{E}T_j^{(H)}$ ($H > 1/2$) and $\mathbf{E}T_j^{(1/2)}$ for $j > 2$ will likely also involve the application of Lemma 1 and Proposition 1.

Acknowledgements

This study was supported by the program for fundamental scientific research of the Siberian Branch of the Russian Academy of Sciences, project no. FWNF-2024-0001.

References

[1] Bahadur, R., 1960. On the number of distinct values in a large sample from an infinite discrete distribution. Proceedings of the National Institute of Sciences of India 26 (A), 67–75.

[2] Karlin, S., 1967. Central Limit Theorems for Certain Infinite Urn Schemes. Journal of Mathematics and Mechanics 17 (4), 373–401.

[3] Samorodnitsky, G., Taqqu, M., 1994. Stable Non-Gaussian Random Processes, Chapman & Hall, New York.

[4] Arkashov, N.S., 2022. On the modeling of stationary sequences using the inverse distribution function. Sib. Electron. Math. Rep. 19 (2), 502–516.

[5] Beran, J., 1994. Statistics for Long-Memory Processes, Chapman & Hall, New York.

[6] Granger, C.W.J. and Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1 (1), 15–29.

[7] Slepian, D., 1962. The one sided barrier problem for Gaussian noise, Bell. Syst. Tech. J. 41, 463–501.

[8] Lindgren, G., Rootzen, H., Leadbetter, M. R., 1983. Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, New York, Berlin, Heidelberg.

[9] Tong, Y.L., 1990. The Multivariate Normal Distribution, Springer-Verlag, New York.

[10] Li, W., Shao, QM., 2002. A normal comparison inequality and its applications, Probab. Theory Relat Fields 122, 494–508.