International Journal of Mathematical Analysis and Applications
2015; 2(1): 1-8
Published online February 10, 2015 (http://www.aascit.org/journal/ijmaa)
ISSN: 2375-3927

American Association for Science and Technology

Keywords
Fractional Brownian Motion,
Estimation of Hurst Parameter,
Unbalanced Sampling,
Maximum Likelihood Estimator,
Fisher Test,
Student's t- Test

Received: January 17, 2015
Revised: February 2, 2015
Accepted: February 3, 2015

A Statistical Study to Provide Estimators of Hurst Parameter for a Fractional Brownian Motion Through Unbalanced Sampling Time

Said Taan El Hajjar

Department of Mathematical Sciences, Ahlia University, Manama, Kingdom of Bahrain

Email address
salhajjar@ahlia.edu.bh

Citation
Said Taan El Hajjar. A Statistical Study to Provide Estimators of Hurst Parameter for a Fractional Brownian Motion Through Unbalanced Sampling Time. *International Journal of Mathematical Analysis and Applications*. Vol. 2, No. 1, 2015, pp. 1-8.

Abstract
This study aims to learn an estimation of the Hurst Parameter for unevenly sampled fractional Brownian motion. The motions are reproduced by means of Cholesky's algorithm, and the parameter of Hurst is predicted by maximizing likelihoods. These methods are proven to be suitable for the use of quantitative data utilised in this paper during simulation. Several tables that contain the estimates of the self-similarity measure are presented in this study according to various sampling procedures with various sizes of motions. The initiations of these tables is stood on a sequence of statistical tests based on Student's *t-test* and Fisher's Test that make it possible to analyze and compare the distinctions between the considered sampling processes. This paper deals with the simulation of the fractional Brownian motions, with their identifications, and with the analysis of the experimental results. This study proves that unpredictable sampling gives more inconsistency between the outcomes and those expected values for a balance sampling. This inconsistency has a tendency to be decreased if there is increase in the size of the signals. Also, this study shows that when there is a uniformlyrandom sampling model, outputs from random samplings tend to be similar to those outputs that come from a balance deterministic sampling. It confirms that the best estimators of Hurst's parameter are obtained by maximizing the chance.

1. Introduction

In this work, we are interesting in providing estimators of Hurst parameter h ($0< h <1$) for a fractional Brownian motion along which the sampling time is not balanced. Instead of the original process which is the usual techniques of simulation by assuming a balancing time sampling, we describes three components that deals with the simulation of the fractional Brownian motions, with their identification, and with the analysis of the experimental results. For this, we need to enable controled and adapted techniques of simulation. These techniques are needed to make sure when the significant estimators of h parameter could be obtained. In contrary to the usual technique, the resolution of such estimator is fit in the unbalanced sampling rather than the balancing one. Then the results are going to be compared with a case of balanced sampling.

The objective of this study is to confirm the theoretical results related to the non bias asymptotic character and its convergence. The main objective is to valid the simulation and the estimation techniques through general cases. To accomplish this, quantitative results are going to be provided to identify the accuracy of estimators for distinct values of Hurst parameter.

2. Overview

The fractional Brownian motion (*FBM*), issued works of Mandelbrot and Van Ness (Mandelbrot & Van Ness, 1968), is part of a process generating the Brownian motion. It posseses the property of self-similarity, quantified by the parameter of Hurst h . The *FBM* finds many applications in signal models: Biology (Mandelbrot, 1983), motion analysis (Scafetta, Griffin & West, 2003), economic series (Carbone, Castelli & Stanley, 2004), image processing (Lundahl, Ohley, Kay &Siffert, 1986), Ethernet network (Vidacs & Virtamo, 2000). Practically, it is difficult for a signal to possess self-similarity. Certain tests are given to verify the best adequate model (Jennane, Harba & Jacquet, 1996); however, these tests are not applicable for the case of signals obtained by unbalanced sampling. The modeling with an *FBM* for this data type can then become inadequate. Most of identification and estimation methods of h for an *FBM* assume an unbalanced sampling [Coeurjolly, (2000); Istas & Lang, (1994); Peltier & Levy-Vehel, (1994)]. These methods require, otherwise, a fundamental change in the algorithms.

The unbalanced sampling of time series gave rise to much work, especially in spectral analysis of unevenly spaced data (Potts, Steidl G. & Tasche, 2000). In this study, we propose an experimental step for *FBM* modeling of an unbalanced signal in times. It is on the one hand to compare the results obtained with the previous studies done over balanced sampling (Jennane, Harba, & Jacquet, 2001), and secondly to determine the possible presence for a bias of the estimator. The estimation of h is the maximum likelihood [(Lehmann, (1980); Lundahl, Ohley, Kay & Siffert, (1986)] who has previously adopted the model proposed by Norros (Norros,1994).

The second section is dedicated to describe *FBM*. A simulation method of unbalanced motions is presented in the third section. We describe in the fourth section the algorithm with maximum likelihood estimator, estimating the parameter h and taking into account the unbalancing of sampling. Before conclusion, we present in the fifth section the experimental results with a set of estimates of h obtained by simulating motions with different kinds of unbalanced timing. We compare the obtained results and assess the presence of bias in the estimation.

3. Description of the fractional Brownian Motion

A *FBM* of Hurst parameter $h \in]\,0\,,1\,[$, noted $[Y(x)]_{x\in\Re}$, is a centered real Gaussian process verified by:

$$Y(0) = 0 \text{ and for every } (x, r) \in \Re^2,\ F\left(|Y(x) - Y(r)|^2\right) = |x - r|^2 \qquad \text{(i)}$$

We deduced the expression of the covariance function defined by:

For every $(x, r) \in \Re^2$,

$$F[Y(x)Y(r)] = 1/2\left(|x|^{2h} + |r|^{2h} - |x-r|^{2h}\right) \qquad \text{(ii)}$$

The process $[Y(x)]_{x\in\Re}$ is self-similar of order h and its motionless mean values:

$$\text{For every } (x,k) \in \Re\times\Re^{+*},\ Y(kx) \sim k^h\, Y(x) \qquad \text{(iii)}$$

$$\text{For every } (x,r) \in \Re^2,\ Y(x) - Y(r) \sim Y(x-r) \qquad \text{(iv)}$$

$[Y(x)]_{x\in\Re}$ is a unique centered continuous Gaussian process self-similar with motionless mean values (Herbin, 2004). In particular, for h = *1/2* , we found the classic Brownian motion where the mean values are more independents. The spectral density $f\ (s)$ of *FBM* mean values behaves as $/s/^{1-2h}$ when s approaches 0, which is a remarkable characteristic of process with long-memory($1/f$)(Beran,1994).

Let λ be the nucleus of fractional integretion of $\tau^2(\Re)$ defined by :

For every $(x,\beta) \in \Re^2$,

$$\lambda(x,\beta) = |x-\beta|^{h-\frac{1}{2}} - |x|^{h-\frac{1}{2}} \qquad \text{(v)}$$

The *FBM* has a representation called moving average. Each variable in the process $Y(x)$ for all x, x is fixed, can be written as a sum of Gaussian reduced centered weighted by the coefficients of the nucleous λ :

For every $x \in \Re$,

$$Y(x) = \int \lambda(x,\beta)\ dS(\beta) \sim N\left(0, \|\lambda(x)\|^2\right) \qquad \text{(vi)}$$

Where S is a Brownian measure, that is an isometry of $\tau^2(\Re)$ tends to a Gaussian Space. The process $(dS(\beta))_{\beta\in\Re}$ is also called white noise. In case of a discrete processes cases, the variables are also Gaussian and each motion is obtained by summing a realization of a Gaussian with a particular weighting, This will be seen in the next section with the Cholesky's algorithm.

4. Simulation Method of Unbalanced Motions

There are numerous methods for simulation of fractional Brownian motions: Stochastic approaches, multi-scale, spectral, and the covariance matrix of the process. Most of these methods assume that the sampling period is being constant. We chose to simulate the motions through the covariance matrix of process using the Cholesky's method, which takes into account the unbalanced of the sampling via equation *(ii)*. The waveletbased methods, those type of Paxon or spectral simulations, provide good simulations, may be used. A simple way would be to simulate a balanced motion then resampling with a non-constant step.

Cholesky's algorithm is formulated as follows: Let ($x_2,\ldots,x_N$) be the sampling instants of motion with length *N-1* and h its Hurst parameter. The particularity of *FBM* is that the covariance matrix ofprocess is completely determined by $x_i$and *h* . We calculate the associated covariance matrixvia equation *(ii)*:

$$\delta_h(i,j) = F(Y(x_i)\,Y(x_j)) =$$

$$\frac{1}{2}(|x_i|^{2h}+|x_j|^{2h}-|x_i-x_j|^{2h})\ with\ 2\le i,j\le N \qquad \text{(vii)}$$

The matrix δ_h is a defined positive symmetric matrix and may, therefore, be divided by the method of Cholesky:

$$\delta_h = AA' \qquad \text{(viii)}$$

Where *A* is the Lower triangular matrix , specifying in the decomposition of Cholesky. If*y*is a sample of length *N - 1* of a Gaussian variable, then the set *(0, Ay)* is a path of length *N* of*FBM* at times$(x_1, \ldots, x_N)$ with $x_1 = 0$ and auto- similarity*h*.

In general, we applied the method of Cholesky to the covariance matrix, noted*B* , of the dirived process at a certain resolution, called *FGN* for Gaussian fractional noise. The method of Cholesky applied to *B* is called exact because the calculated covariance matrix in the discrete case coincide with the continuous case.

In addition, to gain time of calculation, an alternative is to opt for the Levinson's algorithm (*Peltier,1998*) which allows you to reconstruct the array in the decomposition of Choleskythrough the elements of the first row of *B* . In the case of an unbalanced sampling, this algorithm is no longer applicable since *B* is no longer necessarily a matrix of Toeplitz.The unbalancing of stimulous samples is obtained by controling the time values of the parameter *x* which constitutes the basis of the calculation of the covariance matrix of δ_h.

5. Estimation of the Auto-Similarity

5.1. Theoretical Model

The proposed model by Norros*(Norros, 1994)* is used for the study of Ethernet motion,expressing an *FBMY* with a linear tendency for estimating a posteriority which is nottaken into account due to the use of the simulation's method. Without this tendency, the model is expressed in the following manner:

$$T(x) = \sqrt{a}\,Y(x) \qquad \text{(ix)}$$

The process *Y*is a pure Brownian fractional motion, i.e. a process of Gaussian auto-similarity*h* , of mean zero and of covariance matrix δ_h. The quantity"a"is the variance of the variable *T(1)*. In this case, *T* is an *FBM* of auto-similarity *h* and of covariance equals to:

$$F(T(x_i)T(xj)) = a\,F(Y(x_i)Y(x_j)) = a\ \delta_h(i,j) \qquad \text{(x)}$$

Finally we have two parameters to be considered (*a ,h)*which are estimatedby the maximum likelihoodestimator.

5.2. Maximum Likelihood Estimator (*MLE*)

This estimator was adapted by Dahlhaus (Dahlhaus, 1989) for calculating the coefficient of auto-similarity of an*FBM*.The interest of this method is to get benefit of good asymptotic properties. In particular, the estimator $\hat{h}_{MLE}$ converges quite certain to the theoretical value *h* as *N*increases.The choice of *MLE* is motivated by the fact that it does not need to know at prioritythe spectral strength density*(SSD)* of the signal, in contrary to the Whittle estimator. Among the estimation methods of *(SSD),* the simplest is the periodogram method.For an arbitrary sample rate, Scargle has proposed a general version, which has the same statistical properties as the classical periodogram. This estinmator is asymptotically non bias but it is necessary to run with techniques of local smoothing in order to decrease its variance intrinsically high. Cosequently, the *SSD* is estimated whatever the sample rate is, but the approximation of Whittle requires to know the periodogram of an ascending motion, which is not observed directly in the case of non uniformly sampling.

The estimator of the maximum likelihood is formulated as follows: Let $t = (t_1,\ldots,t_N)'$ be a motion of*T* at sampling rate time$x = (x_1,\ldots,x_N)'$. We look to maximizewith respect to (*a* , *h)* a probability of given likelihood by themultinormal density function :

$$A(t;a,h) = (2\pi)^{-\frac{N}{2}}(a^N|\delta_h|)^{-\frac{1}{2}}\,e^{-\frac{1}{2a}t'\delta_h^{-1}t} \qquad \text{(xi)}$$

Where "||" designates the determinant. The logarithm of likelihood is given by:

$$\log A(t;a,h) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(a^N|\delta_h|) - \frac{1}{2a}t'\delta_h^{-1}t \qquad \text{(xii)}$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(a) - \frac{1}{2}\log(|\delta_h|) - \frac{1}{2a}t'\delta_h^{-1}t$$

Maximizing the algorithm of likelihood is to maximize the following quantity:

$$-N\log(a) - \log(|\delta_h|) - \frac{1}{a}t'\delta_h^{-1}t \qquad \text{(xiii)}$$

The estimation of $\hat{a}$ is obtained under the necessary condition that the derivation of equation (xiii) with respect to "a" is null. We obtain:

$$\hat{a} = \frac{1}{N}t'\delta_h^{-1}t \qquad \text{(xiv)}$$

By incorporating the equation (ivx) in (xiii), the problem has at first to estimate h. To near constants, this is resumed to relatively maximize for h the following objective function:

$$-N\log[\ \frac{1}{N}t'\delta_h^{-1}t\ ]-\log|\delta_h| \qquad \text{(xv)}$$

To maximize the objective function, we apply the method quasi-Newton. Strictly, it should demonstrate the concavity of equation (xiii) considered as a function of h. Itis a required difficult question and we would limit it to an experimental verification.

In the case of a randomly sampling, Lai(Lai, 2004) proved that the estimator $\hat{h}$ of Hurst parameter of a *FBM* is a Gaussian asymptote, without bias and standard deviation:

$$\sigma(\hat{h}_{MLE}) = \frac{2}{\det([\ \delta_h^{-1}\delta_h^{'}\ ]^2)} \qquad \text{(xvi)}$$

Where $\delta_h^{'}$ is the derivative matrix of δ_h:

$$\delta_h^{'}(i,j) = \frac{\partial\delta_h(i,j)}{\partial h} \qquad \text{(xvii)}$$

6. Experimental Results

We propose to consider two unbalanced types oftime in the sampling. Through the two types, we proceed by a sub-sampling in a known balanced sample: $\{x_i\}_{i=1}^{n}$ with $x_i - x_{i-1} = 1$. The first sub-sampling denoted by *ICS*is systematic, and it considers alternatively consecutive separating instants of two time units, namely: …., x_i , x_{i+1},x_{i+3}, x_{i+4},….. the second sub-sampling denoted by *ICU* is randomlly chosen and it is obtained through a uniformly selection of a known sample.

6.1. Example

We begin by a simple example proving the diversity of outputs according to sampling. Consider two stimulous signals of size *N = 1024* and with *h = 0.6.* These signals are itterated by Chelosky with respectively a balancesampling rate ($x_i = i$) and an unbalanced issued of the uniform law (*ICU*). These two signals were obtained through the same Gaussian sample (denoted "y" in the third section). We estimate by the Whittle approximation (assumed balanced sampling) and by *MCE*the auto-similarity coefficients of two signals. The outputs are presented in the following table:

Table 1. *Two Signals by Chelosky*

	First Signal (Balanced)	Second Signal (ICU)
Whittle	0.57	0.53
MLE	0.56	0.57

We notice through this example that the estimator of Whittle underestimates the theoretical value of h. The proposed estimator gives a better estimation (tends to 0.6) with a less significant difference. Consequently, The unbalance of the sampling seems to play an important role, which leads us to study its influence in estimating h. This illustrative example does not permit to measure the real efficacy of *MLE* while it is, therefore, for a certain length *N,* to proceed for a performing systematic study by repeating the experimental procedures of simulation and estimation, in a way to study the sampling distribution of the estimator of *h,* its bias and its precision.

6.2. Choosing the Parameters, Strategy of Tests

To lead the analysis, we refer to Jennane and Coll article (Jennane & Harba, 2001). The process is as follows: Having fixed a user sampling, we study signals of different lengths (assume *N*) for different values of the Hurst's exponent (*h*). For a user sampling and for fixed *N* and *h*, we repeat *K* times the operations of simulation of a *FBM* (as the preceding constraints) and of the estimation of *h*. We get, then, a sample of sufficient length to approach the distribution of the estimator h with maximum likelihood. We study then experimentally the bias properties and the convergence of the estimator in question. The choice of *K is* based on controlling the risks , the power of tests, comparing means and variances, which we would achieve. According to the Jennane and Harba, we adopt for the calculation of *K,* the hypothesis along which the variance of the estimator would be equal to the bound of Cramer-Rao correspondence, whereas their article has given the value. This bound depends on *h*, and it adopts a strategy "pessimistic "by choosing the maximum value in all *h* confounded for a fixed *N.* We assume that the density of the estimator is to be Gaussian and with a significance level of *1* %, we find the following table:

Table 2. *signals of different length (assume* N*) with K times repetition of the operations to simulate a FBM*

N	33	65	129	257	513	1025
K	3054	1493	751	384	188	97

In the following, we compare the outputs obtained for each unbalanced sampling with the classical estimation (balancing) and we study the effect of unbalanced time on the bias of estimator. The whole process is summarized in Figure 1.

Before we have validated the observations, we proceed to test the comparisons of samples by counting on the precision and the eventual appear of bias; this for the two unbalanced sampling. In all cases, we consider a risk of significant level α equal to *0.01*. We apply successively Fisher and Student1 t- test (Figure1), in the case of Gaussian, the Fisher's test for comparing variances and then the Student's *t-* test for comparing means. The null hypothesis H_0 of Fisher test is the homogeneity of population variances, which has issued the two samples; the alternative hypothesis H_A is the heterogeneity of variances. Since these have the same length, the statistic of Fisher at $K - 1$ degree of freedom is simply the ratio between s_1^2 and s_2^2, empirical estimators of variances:

$$F(K-1, K-1, \alpha) = \frac{s_1^2}{s_2^2} \text{ where } s_1^2 > s_2^2 \tag{xviii}$$

For a given unbalanced sampling , this test is applicable for each of the counterparts' cases of the two tables. In case where there is no rejection of H_0, we apply the student's *t*- test where the null hypothesis is the equality of the two means of X_1 and X_2. The statistics of the test is as follows:

$$T_{2K-2} = \sqrt{\frac{K-1}{s_1^2 + s_2^2}}\left(\overline{X_1} - \overline{X_2}\right) \tag{xix}$$

Where $\overline{X_1}$ and $\overline{X_2}$ are the sample means of the estimators. If the null hypothesis of the two tests is not rejected, we conclude that the unbalancing of the sampling has no influence on the estimator's properties. Before taking a decision of biasing the estimator, we calculate the confidence interval estimate *(C.I.)* at threshold α Student 2*t*- test (Figure 1) and verifying that it contains the theoretical value *h* of Hurst parameter. An estimation of *h* being independent of others is applying the Central Limit Theorem over $\overline{X}$ and deduces that $\overline{X}$ follows approximately a Gaussian law. Since the standard deviation is unknown, the statistic $\frac{\overline{X}-X}{s/\sqrt{K-1}}$ of sample mean $\overline{X}$ and with sample standard deviation *s* follow approximately a Student's *t*- test law T_{K-1}at$K-1$degree of freedom, which gives the following framework:

$$-t_{\alpha/2} < \frac{\overline{X}-X}{s/\sqrt{K-1}} < t_{\alpha/2} \tag{xx}$$

Where $t_{\alpha/2}$ is (1 – α) / 2 percentile of T_{K-1} . We deduce the confidence interval:

$$\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{K-1}} < X < \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{K-1}} \tag{xxi}$$

In this case, we obtain the *C.I.* that permits to verify whether the chosen theoretical value of the simulation belongs to this interval of confidence or not. The calculation of the *C.I.* is equivalent to testing the null hypothesis $\hat{X} = X$ with the same statistic $\frac{\overline{X}-X}{s/\sqrt{K-1}}$.

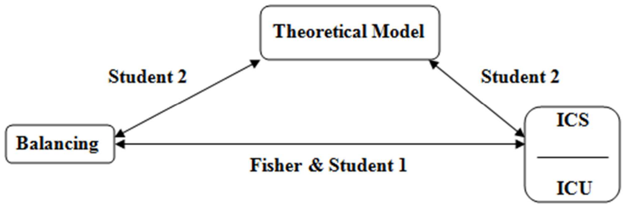

Fig. 1. *Strategic tests for the comparison and the study of bias of the estimator of H by MLE. Student 2 : Student's t- test (interval of confidence);Fisher & Student1 : Fisher test with Student's t- test (comparison of samples)*

6.3. Comparison of Samples (Fisher & Student 1)

The set of calculations was done through the statistical software *PHStat 2*. Table 1 includes the means and the standard deviations obtained through a balanced sampling and crosses the values of *h* with the length of signals. The values in the tables are round to two decimal places but the tests are lead through the exact values. We found the usual properties of the estimation method: The standard deviation of the estimator decreases as *N* increases, illustrating then the fact that *MLE* is convergent. Globally, the means obtained underestimate the values of *h*.

Table 3. *Estimation by MLE with Balancing Sampling*

h \ N	32	64	128	256	512	1024
0.1	0.11±0.07	0.10±0.05	0.10±0.03	0.10±0.02	0.10±0.02	0.10±0.01
0.2	0.20±0.09	0.20±0.06	0.20±0.04	0.20±0.03	0.20±0.02	0.20±0.01
0.3	0.29±0.11	0.30±0.07	0.30±0.05	0.30±0.03	0.30±0.03	0.30±0.02
0.4	0.39±0.12	0.39±0.08	0.39±0.06	0.40±0.04	0.40±0.03	0.40±0.02
0.5	0.49±0.12	0.49±0.08	0.50±0.06	0.50±0.04	0.50±0.03	0.50±0.02
0.6	0.58±0.12	0.59±0.08	0.60±0.06	0.60±0.04	0.60±0.03	0.60±0.02
0.7	0.68±0.12	0.69±0.08	0.69±0.06	0.69±0.04	0.70±0.03	0.70±0.02
0.8	0.77±0.11	0.79±0.08	0.80±0.06	0.80±0.04	0.80±0.03	0.80±0.02
0.9	0.86±0.10	0.88±0.07	0.89±0.05	0.89±0.04	0.90±0.03	0.90±0.02

Table 3 describes the same estimations by taking into account the case of *ICS*. All the standard deviations obtained are less than or equal to that of Table 3. The unbalancing *ICS* reduces, then, the standard deviations of the estimator where the theoretical values of *h* are underestimated. We compared these results with the results obtained in Table 3through more described statistical tests. The shaded entries correspond to a non-rejection of H_0 for Fisher &Student tests. The non-rejection H_0 are comparable for the two tables through the size *N= 512*. For the maximal size *1024,* the results obtained in tables*3* and *4* are similar.

Table 4. *Estimations by MLE with ICS, comparison with Table 3. The shaded entries correspond to a non-rejection of H_0 for Fisher & Student tests.*

h \ N	32	64	128	256	512	1024
0.1	0.11±0.07	0.10±0.05	0.10±0.03	0.10±0.02	0.10±0.02	0.10±0.01
0.2	0.20±0.09	0.20±0.06	0.20±0.04	0.20±0.03	0.20±0.02	0.20±0.01
0.3	0.29±0.10	0.30±0.06	0.30±0.05	0.30±0.03	0.30±0.03	0.30±0.02
0.4	0.39±0.10	0.39±0.07	0.39±0.05	0.40±0.03	0.40±0.02	0.40±0.02
0.5	0.48±0.11	0.49±0.07	0.50±0.05	0.50±0.04	0.50±0.03	0.50±0.02
0.6	0.59±0.11	0.59±0.07	0.60±0.05	0.60±0.04	0.60±0.03	0.60±0.02
0.7	0.68±0.11	0.69±0.08	0.69±0.05	0.69±0.04	0.70±0.03	0.70±0.02
0.8	0.77±0.10	0.79±0.07	0.80±0.05	0.80±0.04	0.80±0.03	0.80±0.02
0.9	0.86±0.09	0.88±0.06	0.89±0.05	0.89±0.04	0.90±0.03	0.90±0.02

Table 5 includes the estimations with unbalanced *ICU*. The results obtained are clearly better than the previous cases in terms of variance. All standard deviations are less than or equal to the results obtained in the previous two tables, and that the difference is more globally powerful for the values of *h* (around 0.5). The non-rejected hypothesis are equivalent in tables 3 and 4 through a size between 512 and 1024. Only the means and the standard deviations for the samples of size 1024 approach the results obtained with a balancing sampling. We recognize that for a periodical unbalancing *ICS*, the estimated values are nearest to the balancing case, the tests of comparing samples are being more rejected in Table 5. In conclusion, the more the sampling is unpredictable, the more the convergence of estimators is slow and this particularity tends to diminish with the size of signals. The sampling does not influence significantly the estimations for signals of low length.

Table 5. *Estimations by MLE with ICU, Comparison with Table 3.*

h \ N	32	64	128	256	512	1024
0.1	0.11±0.06	0.10±0.04	0.09±0.02	0.09±0.01	0.09±0.02	0.10±0.01
0.2	0.20±0.08	0.20±0.05	0.10±0.03	0.10±0.02	0.10±0.02	0.20±0.01
0.3	0.30±0.09	0.30±0.06	0.30±0.03	0.20±0.02	0.20±0.02	0.30±0.02
0.4	0.40±0.08	0.40±0.06	0.40±0.04	0.40±0.03	0.40±0.03	0.40±0.02
0.5	0.50±0.09	0.50±0.06	0.50±0.04	0.50±0.03	0.50±0.03	0.50±0.02
0.6	0.60±0.10	0.59±0.06	0.60±0.04	0.60±0.03	0.50±0.03	0.60±0.02
0.7	0.69±0.10	0.70±0.06	0.70±0.05	0.70±0.03	0.79±0.03	0.70±0.02
0.8	0.78±0.09	0.79±0.06	0.80±0.05	0.80±0.03	0.80±0.03	0.80±0.02
0.9	0.89±0.08	0.89±0.05	0.90±0.05	0.90±0.03	0.90±0.03	0.90±0.02

6.4. Influence of Unbalanced Sampling over the Bias of the Estimator (Student 2)

In this section, we study the importance of the bias of the estimator in terms of the selection of an unbalancing sampling. For this and in complementarities with the calculation of *C.I.,* we apply a Student's *t-* test of the expected value of the estimator with the theoretical value:H_0: $E(\hat{h})=h$. The tables 6, 7and 8 describe the set of *p-values* test for each type of sampling. Since the test is symmetrically two-sided, we should compare the results obtained with the half of the level of significance initially chosen *(α / 2 = 0.005)*. The student's *t-*test is equivalent to determine if the theoretical value belongs to the *C.I.* but at the advantage of specifying the intermediate analysis of *p-values.* For the three tables, the non-rejection of the test corresponds globally to signals with large length. The more the length of signals is large, the more the theoretical value belongs to the *C.I.* whereas the estimator is anon-bias asymptotic. There are*35*non-rejections for the balance of sampling, *30* for *ICS*, *42* for *ICU* and we obtain then from this point of view the best results with *ICU*. The mean of *p-values* of non-rejections is of *0.22* for the balanced sampling, *0.13* for *ICS* and *0.17* for *ICU*. The balanced sampling accepts fewer tests as *ICU*, but the *p-values* are globally higher. In a fixed interval, the more the number of observations is high, the more the sampling *ICU* tends to a balanced sampling. This explains that the results of *ICU* approach the results of balanced sampling in terms of the *p-values* as *N* increases.

Table 6. *Student's t- test H_0: $E(\hat{h})=h$. Set of p-values for the balanced case. The shaded entries correspond to non-rejection of the test (p-values> α / 2).*

h \ N	32	64	128	256	512	1024
0.1	$<0.1\times10^{-15}$	0.43	0.11	0.23	0.07	0.08
0.2	0.17	0.24	0.41	0.25	0.27	0.10
0.3	0.24×10^{-5}	0.005	0.11	0.25	0.08	0.14
0.4	0.52×10^{-11}	0.33×10^{-5}	0.003	0.07	0.38	0.35
0.5	0.16×10^{-12}	0.001	0.09	0.39	0.18	0.23
0.6	$< 0.1\times10^{-15}$	0.52×10^{-4}	0.05	0.40	0.36	0.25
0.7	$< 0.1\times10^{-15}$	0.55×10^{-12}	0.33×10^{-5}	0.002	0.32	0.26
0.8	$< 0.1\times10^{-15}$	0.44×10^{-8}	0.006	0.22	0.41	0.12
0.9	$< 0.1\times10^{-15}$	$< 0.1\times10^{-15}$	0.26×10^{-12}	0.86×10^{-5}	0.28	0.43

Table 7. *Student's t- test H_0: $E(\hat{h})=h$. Set of p-values for the ICS case.*

h \ N	32	64	128	256	512	1024
0.1	$<0.1\times10^{-15}$	0.004	0.06	0.05	0.02	0.05
0.2	0.02	0.48×10^{-3}	0.004	0.02	0.03	0.05
0.3	0.70×10^{-4}	0.01	0.20	0.44	0.46	0.43
0.4	0.23×10^{-6}	0.001	0.06	0.08	0.14	0.14
0.5	0.11×10^{-15}	0.28×10^{-6}	0.10×10^{-3}	0.003	0.01	0.05
0.6	0.94×10^{-12}	0.54×10^{-4}	0.10	0.36	0.32	0.13
0.7	$<0.1\times10^{-15}$	0.88×10^{-6}	0.39×10^{-3}	0.001	0.12	0.08
0.8	$<0.1\times10^{-15}$	0.11×10^{-12}	0.28×10^{-3}	0.07	0.24	0.08
0.9	$<0.1\times10^{-15}$	$<0.1\times10^{-15}$	0.12×10^{-10}	0.004	0.10	0.07

Table 8. *Student's t-test H_0 : $E(\hat{h})=h$. Set of p-values for the ICU case.*

h \ N	32	64	128	256	512	1024
0.1	0.44×10^{-15}	0.26	0.23	0.11	0.15	0.13
0.2	0.003	0.05	0.05	0.06	0.12	0.15
0.3	0.43	0.44	0.49	0.08	0.44	0.17
0.4	0.005	0.09	0.03	0.16	0.04	0.18
0.5	0.002	0.01	0.07	0.38	0.27	0.19
0.6	0.44×10^{-10}	0.67×10^{-4}	0.01	0.23	0.13	0.20
0.7	0.21×10^{-5}	0.10	0.003	0.32	0.06	0.21
0.8	$<0.1\times10^{-15}$	0.27×10^{-8}	0.02	0.27	0.08	0.22
0.9	$<0.1\times10^{-15}$	$<0.1\times10^{-15}$	0.02	0.11	0.03	0.25

6.5. Analysis of *p-Values*

Analyzing the maximum of *p-values* by column, we observe a certain structure in each of the above tables. For the balanced sampling, the maximum is attained for$h = 0.1$, for signals of low lengths, untill$h = 0.9$ for long signals. For *ICS*, the maximum is attained for $h = 0.3$ whatever N,and for *ICU*the dynamics is more randomly but also concerns for most cases of$h = 0.3$. Then theunbalanced sampling has less influence for all h in the neighberhood of *0.1* because the variations of signals are more high. Consequently, the unbalanced *ICS*globally added a bias. For *ICU*, the values obtained through uniformly unbalancing are closely to the theoretical model: The Student 's t - test for *C.I.* accepts *42* values against *30* for *ICS* and *35* for balanced sampling.

So we conclude a reduced bias for a balanced sampling as it is also for *ICU*. The unbalancing" determinist"*ICS* gives significantly poorer results, which is coincided with the results of the *C.I.,* involving a bias more pronounced.

7. Conclusion

We fixed the objective to conduct an experimental study over the estimation of Hurst parameter for fractional Brownian motions along which the sampling time is not balanced. Our study then describes three components. The first component deals with the simulation of the fractional Brownian motions, the second component deals with their identifications, and the third deals with the analysis of the experimental results. The usual techniques of simulation assume a balancing time sampling, which are added to used matrices in the calculation of particular forms that are not setting to unbalanced sampling. The techniques of simulation were able to be controled and adapted. The best estimators for h parameter are obtained by maximizing the chance. In the scientific litterature of the subject, the resolution and the calculation of such estimators assume once again a balanced sampling. We fit the classical techniques in the unbalanced case while we limit the combination. We have intentionally taken the experimental protocols in the case of balanced sampling, proposed by Vidacs & Verdamo (2000) and Jennane, Harba & Jacquet (2001), so that our results are directly comparable with them without any methodological bias.

In general, we naturally confirmed the theoretical results related to maximum likelihood estimator, in particular, the non bias asymptotic character and its convergence. The major interest of our work is to allow the validation of the simulation and the estimation techniques in the general cases. We provide quantitative results in the form of confidence intervals, which allow to identify the precision of estimators for different size of signals and different values of Hurst parameter. We have also considered several types of unbalanced sampling, one systematic with a periodical sampling, the other is randomly chosen with uniformly random sampling. This final analysis confirms in particularthe greatest sensitivity (higher bias) of the maximum likelihood estimation in the symmetrical case with respect to the random method.

References

[1] Beran J., (1994). *Statistics for Long-Memory Processes*. Chapman and Hall New York.

[2] Carbone A., Castelli G. & Stanley, H. (2004). *Time-Dependent Hurst Exponent in Financial Time Series*. Physica A 344. 267-271.

[3] Coeurjolly J. (2000). Simulation and Identification of the Brownian Motion: A Bibliographical and Comparative Study. *Journal of Statistical Software*, 5, 1-53.

[4] DAHLHAUS R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4): 1749–1766.

[5] Herbin E., (2004). *Processus multi-fractionnaires a paramétre multidimensionnels et régularité holdérienne*. PhD thesis, Paris XI Orsay.

[6] Istas J. & Lang G., (1994). Variations Quadratiques et Estimation de Léxposant de Holder Local d'un processus Gaussien. *C. R. Acad. Sci. Paris*, 319, 201-206.

[7] Jennane R., Harba R. & Jacquet G. (1996). Estimation de la Qualite des Methods de Synthese du Movement Brownien Fractionnaire. *Revue Traitement du Signal*, 13 (4), 289-302.

[8] Jennane R., Harba R. & Jacquet G. (2001). Méthodes d'analysedu Movement Brownien Fractionnaire: Théorie at resultants comparatives. *Revue Traitement du Signal*, 18 (5-6), 419-436.

[9] LaiD., (2004).Estimating the Hurst effect and its application in monitoring clinical trials. *Computational Statistics and Data Analysis*, 45 : 549–562.

[10] Lehmann E. (1980). Efficient Likelihood Estimators. *The American Statistician*, 34, 233-235.

[11] Lundahl T., Ohley W., Kay S. & Siffert R. (1986). Fractional Brownian Motion: A Maximum Likelihood estimator and its application to image texture. *IEEE Trans. On Medical Imaging*, 5: 152-161.

[12] Mandelbrot B. and Van J. (1968). Fractional Brownian Motions, fractional noises and applications. *SIAM Review*, 10(4), 422-437.

[13] Mandelbrot B. (1983). *The Fractal Geometry of Nature.* Freeman, New York.

[14] Norrros I. (1994). A Storage model with Self-Similar Input. Queuing Systems, 16, 387-396.

[15] Peltier R., & Levy-Vehel J. (1994). A New Method for Estimating the Parameter of Fractional Brownian Motion. *Research report 2396*, *INRIA Rocquencourt*.

[16] PELTIER R. (1998). *Processus stochastiques fractals avec applications en finance*. PhD thesis, Univ. Paris VI.

[17] Potts D., Steidl G. & Tasche M. (2000). *Fast Fourier Transforms for non-equispaced Data: A tutorial, chapter 12, 253-274. Brikhauser.*

[18] Scafetta N., Griffin L., and West B. (2003). *Holder Exponent Spectra for Human Gait*. Physica A 328. 561-583.

[19] Vidacs A. & Virtamo J. (2000). Parameter estimation of geometrically sampled Fractional Brownian Traffic. *INFOCOM,* 3.