Accepted Manuscript

On the performance of Fisher Information Measure and Shannon entropy estimators

Luciano Telesca, Michele Lovallo

PII: S0378-4371(17)30378-3
DOI: http://dx.doi.org/10.1016/j.physa.2017.04.184
Reference: PHYSA 18313

To appear in: *Physica A*

Received date : 1 November 2016
Revised date : 16 January 2017

Please cite this article as: L. Telesca, M. Lovallo, On the performance of Fisher Information Measure and Shannon entropy estimators, *Physica A* (2017), http://dx.doi.org/10.1016/j.physa.2017.04.184

Highlights

1. Two estimation methods of Fisher Information Measure (FIM) and Shannon entropy (SE) are analysed.
2. One is based on discretizing FIM and SE formulae; the other on the kernel-based estimation of the probability density function.
3. FIM (SE) estimated by using the discrete-based approach is approximately constant with σ, but decreases (increases) with the bin number L.
4. FIM (SE) estimated by using the kernel-based approach is very close to the theoretic value for any σ.

On the performance of Fisher Information Measure and Shannon Entropy estimators

Luciano Telesca[1,*] and Michele Lovallo[2]

[1] National Research Council, Institute of Environmental Analysis, C.da S.Loja, 85050 Tito (PZ), Italy, luciano.telesca@imaa.cnr.it

[*] Corresponding author

[2] ARPAB, 85100 Potenza, Italy

Abstract – The performance of two estimators of Fisher Information Measure (*FIM*) and Shannon entropy (*SE*), one based on the discretization of the *FIM* and *SE* formulae (discrete-based approach) and the other based on the kernel-based estimation of the probability density function (PDF) (kernel-based approach) is investigated. The two approaches are employed to estimate the *FIM* and *SE* of Gaussian processes (with different values of σ and size N), whose theoretic *FIM* and *SE* depend on the standard deviation σ. The *FIM* (*SE*) estimated by using the discrete-based approach is approximately constant with σ, but decreases (increases) with the bin number L; in particular, the discrete-based approach furnishes a rather correct estimation of *FIM* (*SE*) for $L \propto \sigma$. Furthermore, for small values of σ, the larger the size N of the series, the smaller the mean relative error; while for large values of σ, the larger the size N of the series, the larger the mean relative error. The *FIM* (*SE*) estimated by using the kernel-based approach is very close to the theoretic value for any σ, and the mean relative error decreases with the increase of the length of the series. Comparing the results obtained using the discrete-based and kernel-based approaches, the estimates of *FIM* and *SE* by using the kernel-based approach are much closer to the theoretic values for any σ and any N and have to be preferred to the discrete-based estimates.

Keywords: Fisher Information Measure; Shannon entropy; estimation

1. Introduction

Recently, the interest in the application of the Fisher Information Measure (*FIM*) and Shannon entropy (*SE*) to the investigation of time series has been growing. Even though both these measures have been well known in the theoretical statistics [1, 2], only in the last decade they were widely used to investigate the temporal properties of time series to gain information about dynamical systems. In particular, these two quantities play a key role in the context of analysis of complexity of systems, where the study of patterns, structures, and correlation of systems and processes is crucial. Complexity, in fact, depending on the different context (algorithmic, geometrical, computational, stochastic, statistical, etc.) and different field of application (dynamical systems, disordered systems, spatial patterns, language, multielectronic systems, cellular automata, neuronal networks, self-organization, DNA analyses, social sciences, etc.) can be quantified in different manners. One of these makes use of the notion of order/organization and disorder/uncertainty, like the Shannon entropy, which is a general measure of randomness or uncertainty of the probability density, and the Fisher information measure, which quantifies the degree of order or organization of a system. Several applications were performed in the field of volcanology [3, 4], seismology [5], geo-electro-magnetism [6], air pollution monitoring [7] or ecology [8], neurology [9], specific atomic models and densities [15] and general quantum-mechanical central potentials [16].focusing mainly on the discrimination between “anomalous” and “regular” patterns in a dynamical system. Therefore, it is evident that a good estimation of *FIM* and *SE* is crucial in order to get information as most reliable as possible about the dynamics of a system.

In this paper we intend to compare the performance of two methods of estimation of *FIM* and *SE* that have been used so far: one approach (hereafter called “discrete”) is based on discretizing the *FIM/SE* formulae [8, 10], and the other (hereafter called “kernel-based”) is based on the Kernel density estimation of the probability density function (*PDF*) of the time series [4, and references therein].

In carrying out such comparison we will apply both methods to randomly generated time series following a-priori known *PDF*; in particular we will investigate the performance of estimation of *FIM* and *SE* for zero-mean normally distributed time series with various values of variance and sample size.

2. The Fisher Information Measure and the Shannon entropy

The theoretical FIM and SE are defined as follows. Let $f(x)$ be the probability density of a signal x, then its FIM is given by

$$FIM = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial x} f(x) \right)^2 \frac{dx}{f(x)}, \quad (1)$$

and its Shannon entropy is defined as (Shannon, 1948):

$$SE = -\int_{-\infty}^{+\infty} f(x) \log f(x) dx . \quad (2)$$

For a normally distributed process, the *FIM* and *SE* are expressed by the following equations, which depend on the standard deviation σ:

$$FIM = \frac{1}{\sigma^2}, \ SE = \frac{1}{2} \ln\left(2\pi e \sigma^2\right). \quad (3)$$

Fig. 1 shows the variation of the theoretical *FIM* and *SE* with σ for a Gaussian process.

2.1 Discrete approach estimation

Such approach (hereafter indicated by the subscript $_d$) is based on the direct discretization of Eqs. (1) and (2). Given the time series $D=\{x(i),\ i=1,\ldots,N\}$ we perform a regular binning of the series' values (with constant bin size) that produces a set of L disjoint intervals I_m, such that $D=\bigcup_{m=1}^{L} I_m$.

The *FIM* and *SE* are then given by:

$$FIM = \sum_{m=1}^{L-1} \frac{\left[p(I_{m+1}) - p(I_m)\right]^2}{p(I_m)} \quad (4)$$

$$SE = -\sum_{m=1}^{L} p(I_m) \log\left(p(I_m)\right) \quad (5)$$

that are valid if $p(I_{m+1})$ and $p(I_m)$ are non-zero. The probability $p(I_m)$ is the ratio between the number of series' values within the interval I_m and the total length N of the series. In such approach, the parameter that has to be selected a priori is the number of bin L.

2.2 Kernel-based approach

Such approach (hereafter indicated by the subscript $_k$) is based on the application of the plug-in method for the Kernel density estimation of the pdf of the series' values

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (6)$$

where $K(x)$ is the kernel estimator, n the size of the series, and h the bandwidth [11, 12]. This method optimizes the bandwidth of the kernel density estimator in order to estimate the pdf f. The optimal value of the bandwidth is given by:

$$h* = n^{-1/5} \left(J(f)\right)^{-1/5} \left(M(K)\right)^{1/5} \quad (7)$$

where

$$J(f) = \int_{-\infty}^{+\infty} \left(f''(x)\right)^2 dx \quad (8)$$

with $f''(x)$ the second derivative of f and

$$M(K) = \int_{-\infty}^{+\infty} K^2(u) du. \quad (9)$$

By using the plug-in method the bandwidth h is approximated by means of an iterative approximation of $J(f)$. Thus, a sequence of positive numbers $h(k)$ is constructed through the iterations, where k indicates the number of iterations.

The steps of the plug-in method are the following:

1) The integration of $M(K)=\int_{-\infty}^{+\infty}K^2(u)du$ is computed, where *K(u)* is the Gaussian kernel;
2) $J^0(f)$ is arbitrarily initialized and *h(0)* is determined by using Eq. (7);
3) The pdf *f(0)* is estimated by using Eq. (6) with *h(0)*;
4) At the *k*-th iteration, $J^k(f)$ is calculate from *f(k-1)* by using Eq. (8);
5) The bandwidth *h(k)* is computed by using Eq. (7) and *f(k)* is re-estimated by means of Eq. (6)
6) The iteration will stop if $\left|\frac{f^{(k-1)}(x_i)-f^{(k)}(x_i)}{f^{(k)}(x_i)}\right|<\varepsilon,\ \forall x_i$
7) Then the *FIM* and *SE* are calculated by using Eqs. (1) and (2) where *f(x)* is substituted by $\hat{f}(x)$

3. Results

The aim of this paper is to compare the performance of the discrete- and kernel-based estimators of *FIM* and *SE*. To this purpose, we generated one hundred time series of random values drawn from zero-mean normal distributions with standard deviation σ ranging from 0.1 to 5.9 (with step of 0.1), and for three different sizes N=10^3, 10^4 and 10^5.

Fig. 2a shows the comparison between the theoretic *FIM* and the $<FIM_d>$ that is calculated averaging one hundred values of FIM_d, which is the *FIM* computed for a random series by using the discrete approach; the results are shown versus the standard deviation σ and correspond to the case N=10^3 and L=5. Fig. 2b shows the comparison between theoretic *SE* and the $<SE_d>$. Both the discrete-based *FIM* and *SE* do not estimate correctly the theoretic value of *FIM* and *SE*. Increasing L, and keeping N=10^3, $<FIM_d>$ and $<SE_d>$ are approximately constant with σ, but change with L; in particular $<FIM_d>$ decreases and $<SE_d>$ increases with L (Fig. 3). Furthermore, $<FIM_d>$ coincides with the theoretic *FIM* for L and σ satisfying an approximate linear relationship L~7.47σ (Fig. 4a), which indicates that in order to estimate correctly the *FIM* of a Gaussian process, larger σ higher L.

Analogously, the $<SE_d>$ (shown in Fig. 3 for several values of L) is a good estimator of the theoretic SE, if L~6.66σ (Fig. 4b).

Even changing the size of the series, $<FIM_d>$ and $<SE_d>$ are almost constant with the standard deviation of the Gaussian process, but change with the bin number L (Fig. 5). Also for larger sizes of the data set the discrete approach furnishes a correct estimation of FIM and SE if L and σ satisfy respectively the approximate linear relationships L~8.28σ and L~7.78σ (N=10^4), and L~9.25σ and L~8.78σ (N=10^5).

The mean relative error is defined as

$$<\varepsilon_{F,d(k)}>=<\left|\frac{FIM_{d(k)}-FIM}{FIM}\right|>,\ <\varepsilon_{S,d(k)}>=<\left|\frac{SE_{d(k)}-SE}{SE}\right|> \quad (10)$$

where the subscript $_d$ or $_k$ refers to the discrete or kernel-based approach. Fig. 6 shows the variation with σ of the mean relative errors in the discrete case for L=5,…,15 and N=10^3: each relative error curve presents a minimum that shifts forward with the increase of σ. The curves of the $<\varepsilon_{S,d}>$ (Fig. 6b) present also a maximum at σ=0.2. Comparing the mean relative error curves for a fixed value of L and different value of the size of the series, we observe that for small values of σ, the larger the size of the series, the smaller the mean relative error; but for large values of σ, the larger the size of the series, the larger the mean relative error (Figs. 7 and 8).

Fig. 9a shows the comparison between the theoretic FIM and the $<FIM_k>$ versus the standard deviation σ and corresponding to the case N=10^3. Fig. 9b shows the comparison between theoretic SE and the $<SE_k>$. Both the kernel-based FIM and SE furnish very close estimates of the theoretic value of FIM and SE for any σ. The mean relative error $<\varepsilon_{F,k}>$ for a fixed size does not change with the σ, but decreases with the increase of the length of the series. The mean relative error $<\varepsilon_{S,k}>$ is maximum for σ=0.2, but decreases with both the increase of σ and the size N.

It has been recently proposed another estimator in particular for the FIM. Sanchez et al. (2009) defined the discrete estimator for FIM as

$$FIM = F_0 \sum_{m=1}^{L-1} \left[\left(p_{m+1}\right)^{\frac{1}{2}} - \left(p_m\right)^{\frac{1}{2}} \right]^2 \quad (11)$$

where F_0 is a normalization constant with value (Goncalves et al., 2016)

$$F_0 = \begin{cases} 1 \quad if \;\; p_{m^*} = 1 \;\; for \;\; i^* = 1 \; or \; i^* = L \;\; and \;\; p_m = 0 \;\; \forall m \neq m \\ \\ 1/2 \quad otherwise \end{cases} \qquad .(12)$$

We evaluated the performance of this particular estimator of *FIM* on the base of one hundred time series of random values drawn from zero-mean normal distributions with standard deviation σ ranging from 0.1 to 5.9 (with step of 0.1), and for size N=10^3. The results are similar to those obtained for the discretization method (Eq. 4). Fig. 12a shows the comparison between the theoretic *FIM* and the $<FIM_d>$ that is calculated averaging one hundred values of FIM_d, which is the *FIM* computed for a random series by using Eq. 11; the results are shown versus the standard deviation σ and correspond to the case N=10^3 and L=5. Fig. 12b shows the mean relative error $<\varepsilon_{F,d}>$ varying with σ: the mean relative error is considerably lower than that obtained by using the discretization formula (Eq. 4) (see Fig. 6a for comparison), however it is larger than that obtained by using the kernel-based approach (see Fig. 10 for comparison) that still represents the best method to estimate the FIM.

4. Conclusions

The Fisher Information Measure (*FIM*) and Shannon entropy (*SE*) are statistical tools used to investigate the evolution of dynamical systems. Although they are used for years and in diverse scientific fields, no study has been carried out on their correct estimation until now. Two approaches have been used for their estimation so far: one based on the simple discretization of their formulae and the other based on the kernel-based estimation of the *PDF*. The present study intends to definitely establish which of both methods performs better, and, so, has to be preferred when *FIM* and *SE* are calculated to get information about the time dynamics of a system.

The calculation of both quantities for series of Gaussian processes with changing standard deviation and size has shown that the kernel-based approach furnishes estimates of *FIM* and *SE* closer to the theoretic values than those obtained by using the discrete-based approach, and so the kernel-based approach has to be preferred.

---

5. References

[1] R.A. Fisher, Math. Proc. Cambridge Philos. Soc. **22**, 700 (1925).

[2] C. E. Shannon, The Bell System Technical Journal 27, 379, 623 (1948).

[3] L. Telesca, M. Lovallo and R. Carniel, J. Volcanol. Geoterm. Res. 195, 78 (2010).

[4] L. Telesca, M. Lovallo, J. Martì Molist, C. López Moreno and R. Abella Meléndez, Terra Nova 26, 425 (2014).

[5] L. Telesca, M. Lovallo, V. Alcaz and I. Ilies, Physica A 421, 541 (2015); P. A. Varotsos, N. V. Sarlis, E. S. Skordas, H. K. Tanaka, and M. S. Lazaridou, Phys. Rev. E 74, 021123 (2006).

[6] L. Telesca, M. Lovallo, H.-L. Hsu and C.-C. Chen, Physica A 390, 1350 (2011).

[7] A. B. Chelani, Stoch Environ Res Risk Assess 28, 921 (2014).

[8] B. D. Fath, H. Cabezas and C. W. Pawlowski, Journal of Theoretical Biology 222, 517 (2003).

[9] F. Montani and O. A. Rosso, Entropy 16, 4677 (2014).

[10] M. T. Martin, F. Pennini and A. Plastino, Phys. Lett. A 256, 173 (1999).

[11] V. C. Raykar and R. Duraiswami, Proceedings of the sixth SIAM International Conference on Data Mining, 524 (2006).

[12] M. Troudi, A. M. Alimi and S. Saoudi, EURASIP Journal on Advances in Signal Processing, Article ID 739082 (2008).

[13] P. Sanchez-Moreno, J. S. Dehesa and R. J. Yañez, Discrete densities and Fisher Information, in Proceedings of the 14th International Conference on Difference Equations and Applications.

Uğur-Bahçeşehir University Publishing Company, Istanbul, Turkey Difference Equations and Applications: 291 (2009).

[14] B. A. Gonçalves, L. Carpi, O. A. Rosso, M. G. Ravetti, Physica A 464, 93 (2016).

[15] A. Nagy and K. D. Sen, Phys. Lett. A 360, 291 (2006)

[16] E. Romera, P. Sánchez-Moreno and J. S. Dehesa, J. Math. Phys. 47, 103504 (2006)

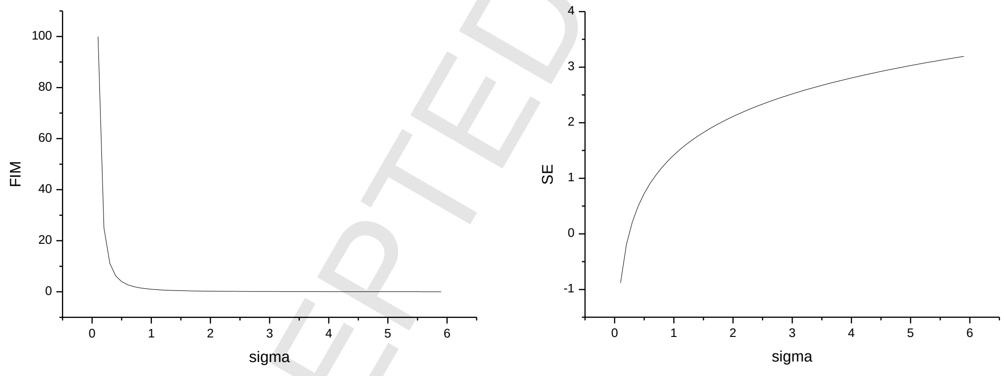
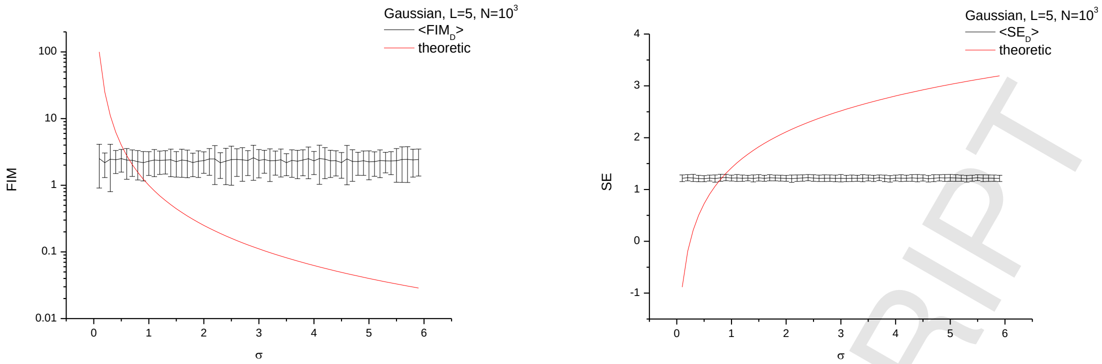


Fig. 1. Variation of the theoretic *FIM* and *SE* for a Gaussian process with σ ranging from 0.1 to 5.9.



a) b)

Fig. 2 Mean (± standard deviation) of the FIM_d (a) and SE_d (b) calculated over 100 random time series normally distributed with size N=10^3, and number of bin L=5. The red curves in both plots represent the theoretic *FIM* (a) and *SE* (a) versus σ.

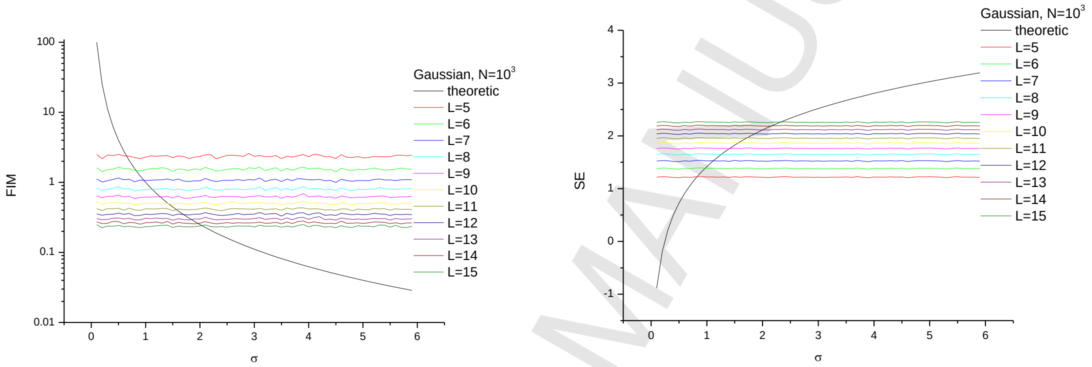
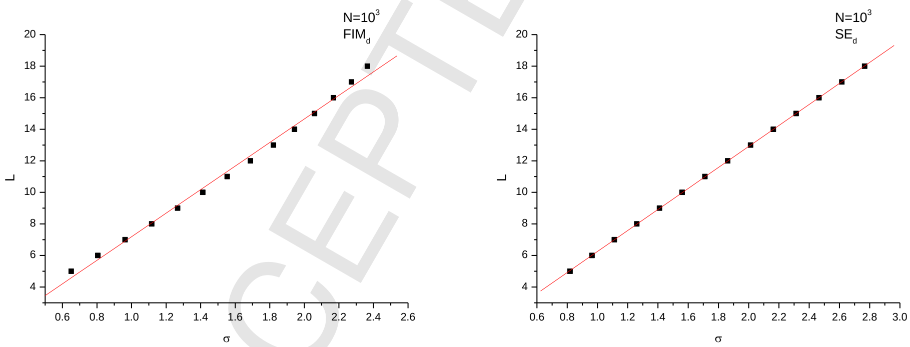


Fig. 3. Mean (± standard deviation) of the FIM_d (a) and SE_d (b) calculated over 100 random time series normally distributed with size N=10^3, and number of bin L from 5 to 15.



a) b)

Fig. 4. Relationship between L and σ for *FIM* (a) and *SE* (b).

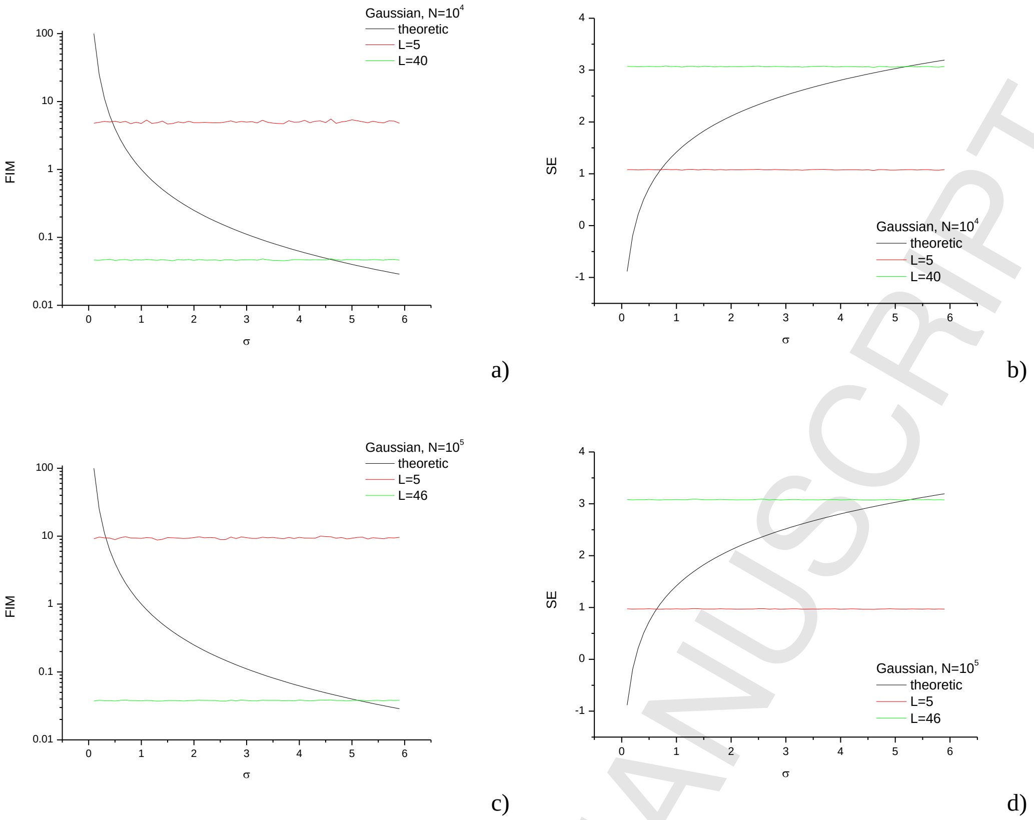


Fig. 5. Mean of the FIM_d and SE_d versus σ calculated over 100 random time series normally distributed with size N=10^4 (a, b) and N=10^5 (c, d).

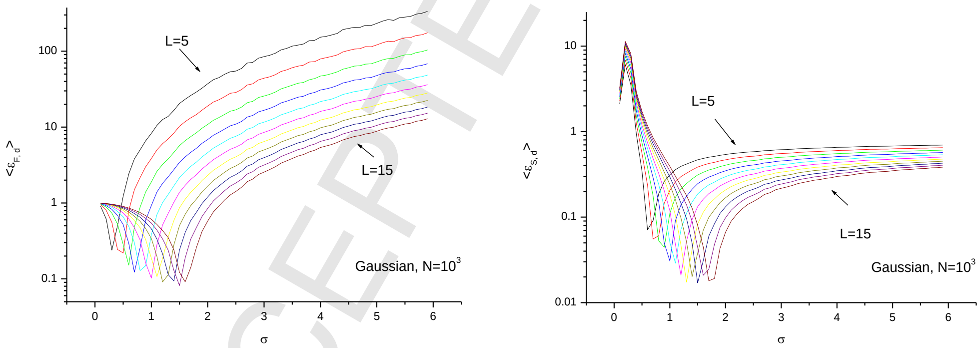


Fig. 6. Mean relative error of *FIM* (a) and *SE* (b) versus σ, in the discrete approach and for L=5,…,15 The size of the series is N=10^3.

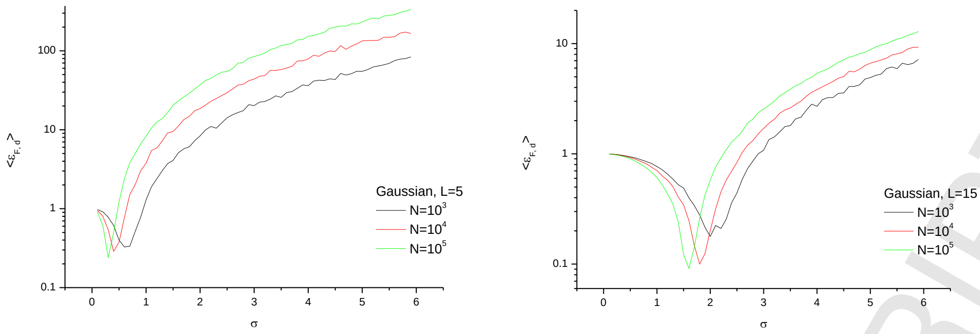


Fig. 7. Mean relative error of *FIM* versus σ, in the discrete approach and for L=5 (a) and L=15 (b) and size of the series is N=10^3, 10^4 and 10^5.

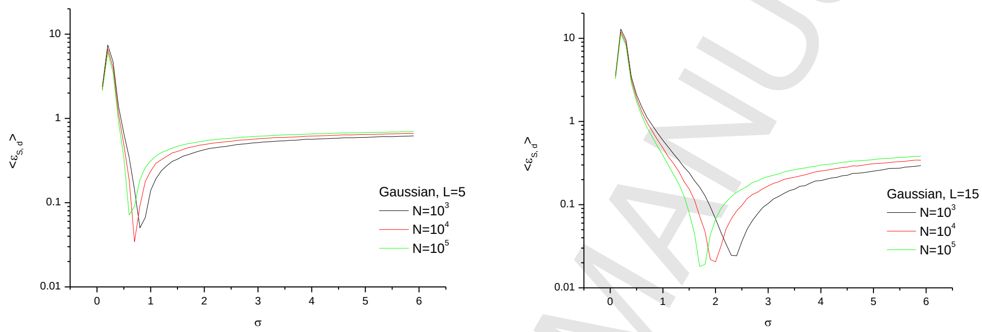


Fig. 8. Mean relative error of *SE* versus σ, in the discrete approach and for L=5 (a) and L=15 (b) and size of the series is N=10^3, 10^4 and 10^5.

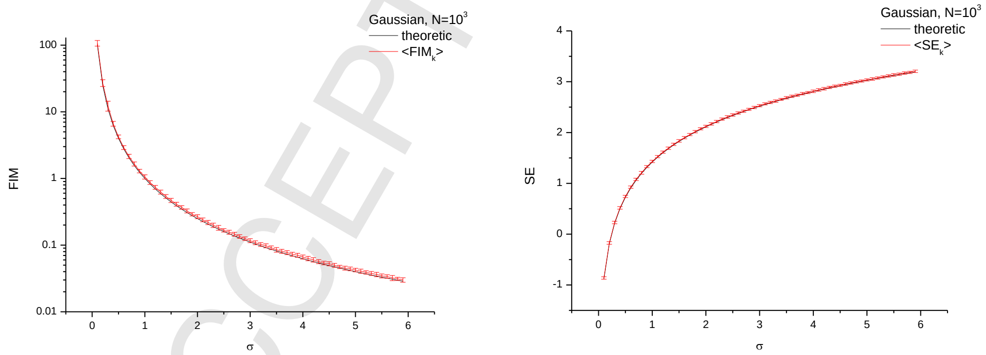


Fig. 9 Mean (± standard deviation) of the FIM_k (a) and SE_k (b) calculated over 100 random time series normally distributed with size N=10^3. The red curves in both plots represent the theoretic *FIM* (a) and *SE* (a) versus σ.

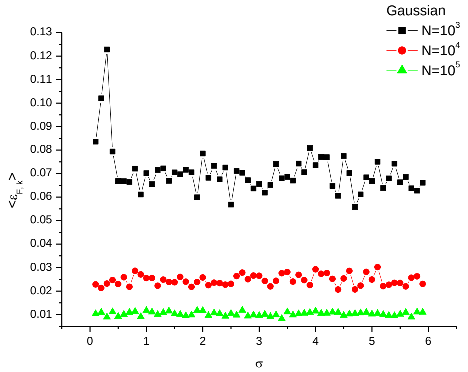


Fig. 10. Mean relative error of *FIM* versus σ, in the kernel-based approach. The size of the series is N=10^3, 10^4 and 10^5.

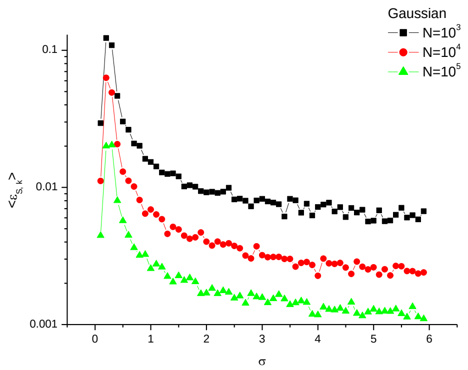
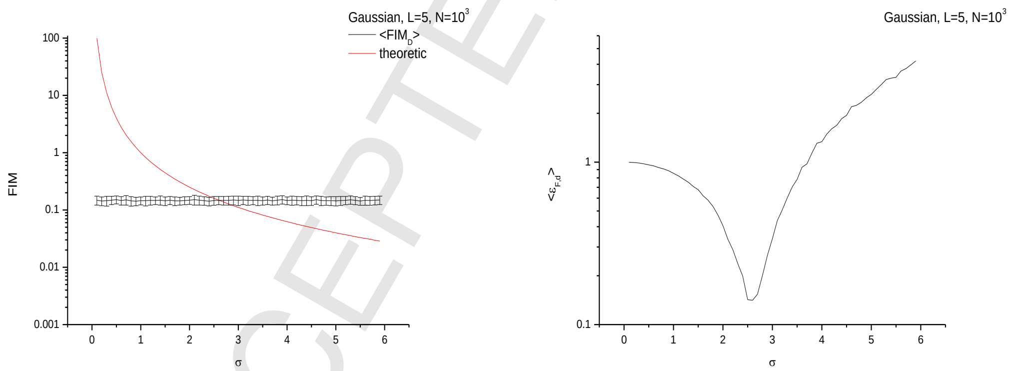


Fig. 11. Mean relative error of *SE* versus σ, in the kernel-based approach. The size of the series is N=10^3, 10^4 and 10^5.



a) b)

Fig. 12 Mean (± standard deviation) of the FIM_d (a) and its mean relative error (b) calculated by using Eq. 11 over 100 random time series normally distributed with size N=10^3, and number of bin L=5. The red curve represents the theoretic *FIM* versus σ.