MDPI

The Systematic Bias of Entropy Calculation in the Multi-Scale Entropy Algorithm

Jue Lu [1,2] and **Ze Wang** [2,*]

1 School of Mathematics, Physics and Information Science, Shaoxing University, Shaoxing 312000, China; admiral_lu@hotmail.com

2 Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA

* Correspondence: ze.wang@som.umaryland.edu

Abstract: Entropy indicates irregularity or randomness of a dynamic system. Over the decades, entropy calculated at different scales of the system through subsampling or coarse graining has been used as a surrogate measure of system complexity. One popular multi-scale entropy analysis is the multi-scale sample entropy (MSE), which calculates entropy through the sample entropy (SampEn) formula at each time scale. SampEn is defined by the "logarithmic likelihood" that a small section (within a window of a length m) of the data "matches" with other sections will still "match" the others if the section window length increases by one. "Match" is defined by a threshold of r times standard deviation of the entire time series. A problem of current MSE algorithm is that SampEn calculations at different scales are based on the same matching threshold defined by the original time series but data standard deviation actually changes with the subsampling scales. Using a fixed threshold will automatically introduce systematic bias to the calculation results. The purpose of this paper is to mathematically present this systematic bias and to provide methods for correcting it. Our work will help the large MSE user community avoiding introducing the bias to their multi-scale SampEn calculation results.

Keywords: entropy; multi-scale sample entropy; systematic bias

1. Introduction

Complexity is an important property of a complex system such as the living organisms, Internet, traffic system etc. Measuring system complexity has long been of great interest in many research fields. Since complexity is still elusive to define, a few approximate metrics have been used to quantify complexity. One widely used measure is entropy which quantifies the irregularity or randomness. Complexity and entropy, however, diverge when complexity reaches the peak. Before the peak, complexity increases with complexity, but complexity decreases with entropy after the peak. To provide approximate solution to this dilemma, people have proposed many empirical measures. A popular one is the multi-scale entropy (MSE) proposed by Costa et al. [1]. MSE is based on Sample entropy (SampEn) [2,3], which is an extension of the well-known Approximate entropy (ApEn) [3,4] after removing the self-matching induced bias. SampEn has gained popularity in many applications such as neurophysiological data analysis [5] and functional MRI data analysis [6,7] because of the relative insensitivity to data length [2,8]. Because complex signal often presents self similarity when the signal is observed at different time scale, Costa et al first applied SampEn to the same signal but at different time scales after coarse graining. When applied to Gaussian noise and 1/f noise, it was observed that SampEn of Gaussian noise decreases with the signal subsampling scale while it stays at the same level for most of scales of a 1/f process. Since a 1/f process is known to have higher complexity (defined by the higher self similarity) than Gaussian noise, the diverging MSE of a 1/f noise and the Gaussian noise appears to support that MSE may provide an approximate

approach to measure system complexity. Since its introduction, MSE has been widely used in many different applications as reflected by the thousands of paper citations [1,9]. While MSE and its variants have been shown to be effective for differentiating different system states through simulation or real data, it introduces bias by using the same threshold for identifying the repeated transit status at all time scales. Nikulin and Brismar [10] first observed that MSE not purely measures entropy but both entropy and variation at different scales. We here claimed that the changing variation captured by MSE is mainly caused by an incomplete scaling during the coarse-graining process and the subsequent variance change induced entropy change should be considered as a systematic bias to be removed.

The rest of this report is organized as follows. Section 2 is the background. To better understand the series of entropy formation, we introduced Shannon entropy, ApEn, SampEn, and MSE. Section 3 describes the bias caused by the coarse-graining process and the one threshold-for-all-scales MSE algorithm. Both a mathematical solution and a practical solution were provided to correct the bias. Section 5 concludes the paper.

2. Entropy and MSE

This section provides a brief history about the evolution of entropy and approximate entropy measures.

Hartley and Nyquist first used logarithm to quantify information [11,12]. Shannon then proposed the concept of Shannon entropy as a measure of information through the sum of the logarithmically weighted probability [13]. Denoting a discrete random variable by X and its probability by $p(x)$, Shannon entropy of X is formulated as:

$$H(X) = - \sum_{x \in X} p(x) \log p(x) = E[\log (\frac{1}{p(x)})];$$

In an analogous manner Shannon defined the entropy of a continuous distribution with the density distribution function(pdf) $p(x)$ by:

$$H(X) = - \int_{x \in X} p(x) \log p(x) \, \mathrm{d}x = E[- \log p(x)],$$

where E represent the expectation operator. Without loss of generality, in this paper we use natural logarithms to calculate entropy. When the entropy calculated via a logarithm to base b, it could be calculated by $H_b(X) = \frac{1}{\log b} H(X)$.

Shannon entropy was then extended into the Kolmogorov–Sinai(K-S) entropy [14] for characterizing a dynamic system. Assume that the F-dimension phase space is partitioned into a collection of cells of size r^F and the state of the system is measured at constant time intervals δ. Let $p(c_1, ...c_n)$ be the joint probability that the state of system $x(t = \delta)$ is in cell c_1, $x(t = 2\delta)$ is in cell c_2, ... , and $x(t = n\delta)$ is in cell c_n. The K-S entropy is defined as

$$\text{K-S entropy} = - \lim_{\delta \to 0} \lim_{r \to 0} \lim_{n \to \infty} \frac{1}{\delta n} \sum_{c_1, ..., c_n} p(c_1, ..., c_n) \log p(c_1, ..., c_n).$$

K-S entropy depends several parameters and is not easy to estimate. To solve this problem, Grassberger and Procaccia [15] proposed K_2 entropy as a lower bound of K-S entropy. Given a time series $U = \{u_1, u_2, ..., u_N\}$ with length N, define a sequence of m dimension vectors $v_i^{(m)} = [u_i, u_{i+1}, ..., u_{i+m-1}]$, $1 \leq i \leq N - m + 1$. The m dependence of functions are

$$C_i^m(r) = (N - m + 1)^{-1} \sum_{j=1}^{N-m+1} \theta(r - \|v_i^{(m)} - v_j^{(m)}\|)$$

and

$$C^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} C_i^m(r)$$

where $\|v_i - v_j\|$ is Euclidean metric $\|v_i - v_j\| = \left(\sum_{h=0}^{m-1}(u_{i+h} - u_{j+h})^2\right)^{\frac{1}{2}}$ and $\theta(\cdot)$ is Heaviside step function. K_2 entropy is defined as

$$K_2 \text{ entropy} = \lim_{r\to 0}\lim_{m\to\infty}\lim_{N\to\infty}\frac{1}{\delta}\log\frac{C^m(r)}{C^{m+1}(r)}.$$

By incorporating the embedding vector based phase space reconstruction idea proposed by Takens [16] and replacing the Euclidean metric with the Chebyshev metric $\|v_i - v_j\| = \max_{h=0}^{m-1}|u_{i+h} - u_{j+h}|$, Eckmann and Ruelle [17] proposed an estimate of the K-S entropy through the so-called E-R entropy:

$$\Phi^m(r) = (N - m + 1)^{-1}\sum_{i=1}^{N-m+1}\log C_i^m(r)$$

$$\text{E-R entropy} = \lim_{r\to 0}\lim_{m\to\infty}\lim_{N\to\infty}\frac{1}{\delta}\left[\Phi^m(r) - \Phi^{m+1}(r)\right],$$

where the delay is often set to be $\delta = 1$.

The E-R entropy has been useful in classifying low-dimensional chaotic systems, but it becomes infinity for a process with superimposed noise of any magnitude [18]. Pincus [4] then extended the E-R entropy into the now well-known ApEn depending on a given embedding window length m and a distance cutoff r for the Heaviside function:

$$ApEn(U; m, r) = \Phi^m(r) - \Phi^{m+1}(r),$$

and

$$ApEn(m, r) = \lim_{N\to\infty} ApEn(U; m, r), \ N \text{ is the length of discrete signal } U.$$

SampEn was proposed by Richman and Moorman [19] as an extension of ApEn to avoid the bias induced by countering the self-matching of each of the embedding vectors. Specifically, SampEn is formulated by:

$$B_i^m(r) = (N - m - 1)^{-1}\sum_{j=1, j\neq i}^{N-m}\theta(r - \|v_i^{(m)} - v_j^{(m)}\|),$$

$$B^m(r) = (N - m)^{-1}\sum_{i=1}^{N-m}B_i^m(r),$$

$$A_i^m(r) = (N - m - 1)^{-1}\sum_{j=1, j\neq i}^{N-m}\theta(r - \|v_i^{(m+1)} - v_j^{(m+1)}\|),$$

$$A^m(r) = (N - m)^{-1}\sum_{i=1}^{N-m}A_i^m(r),$$

$$SampEn(U; m, r) = -\log\frac{A^m(r)}{B^m(r)}, \ \text{fix } m \text{ and } r,$$

$$SampEn(m, r) = \lim_{N\to\infty} SampEn(U; m, r), \ N \text{ is the length of discrete signal } U.$$

The coarse-graining multi-scale entropy-based complexity measurement can be traced back to the work by Zhang [20] and Fogedby [21]. In [1,22] Costa et al. calculated entropy at each coarse-grained scale using SampEn and named this process as the MSE. As commented by Nikulin and Brismar [10], a problem of the MSE algorithm is the use of the same matching criterion r for all scales, which causes systematic bias to SampEn.

3. The Systematic Bias of Entropy Calculation in MSE

In MSE [1,22], the embedding vector matching threshold r in defined by the standard deviation of the original signal. Using the same threshold, entropy of Gaussian signal decreases with the scale used to downsample the original signal. By contrast, entropy of $1/f$ signal remains unchanged when scale increases. As $1/f$ signal is known to have high complexity while Gaussian noise has a very low complexity, the monotonic MSE decaying trend or the sum of MSE at different scales were proposed as a metric for quantifying signal complexity.

However, the moving-average based coarse-graining process automatically scales down the subsampled signal at different time scales. Without correction, this additional multiplicative scaling will be propagated into the standard deviation of the signal to be assessed at each time scale and will artificially change sample entropy. This bias can be easily seen from the coarse-graining of a Gaussian noise.

Denote a Gaussian variable and its observations by $X = \{x_1, x_2, ..., x_N\}$, where N indicates the length of the time series. The coarse-graining or moving averaging process can be described by $Y^{(\tau)} = \{y_j^{(\tau)}\}$, $y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i$ where $\tau > 0$ is the coarse-graining level or the so-called "scale". Given the mutual independence of the individual samples of X, the moving averaging of these samples can be considered as an average of independent random variables rather than observations of a particular random variable. In other word, we can rewrite $Y^{(\tau)}$ to be $Y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} X_i$, where X_i is a random variable. For Gaussian noise X, X_i will be Gaussian noise too and can be fully characterized with the same mean μ and standard deviation (SD) σ. Through a simple mathematics operation, we can get that $\mathrm{SD}(Y^{(\tau)}) = \sigma/\sqrt{\tau}$. Because $\mathrm{SD}(\tau)$ monotonically decreases with τ, if we do not adjust the matching threshold, the number of matched embedded vectors will increase with τ, resulting a decreasing SampEn.

Entropy of a Gaussian distributed variable can be calculated through Shannon entropy:

$$
\begin{aligned}
H(Y) &= -\int_{-\infty}^{+\infty} p(y) \log p(y) \, \mathrm{d}y \\
&= -\int_{-\infty}^{+\infty} p(y) \log(\frac{1}{\sigma_y \sqrt{2\pi}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}) \, \mathrm{d}y \\
&= -\int_{-\infty}^{+\infty} p(y) \log(\frac{1}{\sigma_y \sqrt{2\pi}}) \, \mathrm{d}y - \int_{-\infty}^{+\infty} p(y) \log(e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}) \, \mathrm{d}y \\
&= -\log(\frac{1}{\sigma_y \sqrt{2\pi}}) \int_{-\infty}^{+\infty} p(y) \, \mathrm{d}y + \frac{1}{2\sigma_y^2} \int_{-\infty}^{+\infty} (y-\mu_y)^2 p(y) \, \mathrm{d}y \\
&= \frac{1}{2} \log(2\pi \sigma_y^2) + \frac{1}{2}.
\end{aligned}
$$

For the simplicity of description, we often normalize the random variable to have a $\mu = 0$ and $\sigma = 1$. Considering the scale-dependent SD derived above, we can then get the Shannon entropy of the Gaussian variable at the scale τ by

$$
H(Y^{(\tau)}) = \frac{1}{2} \log\left(\frac{2\pi}{\tau}\right) + \frac{1}{2}
$$

This equation clearly demonstrates the non-linearly but monotonically decreasing relationship of entropy with respect to scale τ.

Below, we provided mathematical derivation of the dependence of MSE on the signal subsampling scale. Given the m dimensional embedding vectors $Z_j^{(m)} = [Y_j, Y_{j+1}, ..., Y_{j+m-1}]$, sample entropy can be expressed as [22]

$$
\begin{aligned}
SampEn(Y; m, r) &= -\log \frac{\Pr(\|Z_j^{(m+1)} - Z_i^{(m+1)}\| \le r)}{\Pr(\|Z_j^{(m)} - Z_i^{(m)}\| \le r)} \\
&= -\log \Pr\left(\|Z_j^{(m+1)} - Z_i^{(m+1)}\| \le r \,\middle|\, \|Z_j^{(m)} - Z_i^{(m)}\| \le r\right).
\end{aligned}
$$

where $\| \cdot \|$ is the Chebyshev distance.

For $m = 1$, we can have

$$
\{\|Z_j^{(m)} - Z_i^{(m)}\| \le r\} = \{|Y_j - Y_i| \le r\},
$$

and

$$
\begin{aligned}
\{\|Z_j^{(m+1)} - Z_i^{(m+1)}\| \le r\} &= \{\max\{|Y_j - Y_i|, |Y_{j+1} - Y_{i+1}|\} \le r\} \\
&= \{|Y_j - Y_i| \le r\} \wedge \{|Y_{j+1} - Y_{i+1}| \le r\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
SampEn(Y; m, r) &= -\log \frac{\Pr(\{|Y_j - Y_i| \le r\} \wedge \{|Y_{j+1} - Y_{i+1}| \le r\})}{\Pr(|Y_j - Y_i| \le r)} \\
&= \Pr\left(|Y_{j+1} - Y_{i+1}| \le r \,\middle|\, |Y_j - Y_i| \le r\right).
\end{aligned}
$$

Based on the iid condition of Y_j, we can draw a conclusion that

$$
\Pr\left(|Y_{j+1} - Y_{i+1}| \le r \,\middle|\, |Y_j - Y_i| \le r\right) = \Pr\left(|Y_{j+1} - Y_{i+1}| \le r\right).
$$

If $m \ge 2$, we can get

$$
\{\|Z_j^{(m)} - Z_i^{(m)}\| \le r\} = \{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\},
$$

and

$$
\begin{aligned}
\{\|Z_j^{(m+1)} - Z_i^{(m+1)}\| \le r\} &= \{\max_{k \in \{0,...,m\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\} \\
&= \{|Y_{j+m} - Y_{i+m}| \le r\} \wedge \{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&SampEn(Y; m, r) \\
&= -\log \frac{\{|Y_{j+m} - Y_{i+m}| \le r\} \wedge \{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\}}{\{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\}} \\
&= \Pr\left(|Y_{j+m} - Y_{i+m}| \le r \,\middle|\, \{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\}\right).
\end{aligned}
$$

and

$$
\Pr\left(|Y_{j+m} - Y_{i+m}| \le r \,\middle|\, \{\max_{k \in \{0,...,m-1\}} \{|Y_{j+k} - Y_{i+k}|\} \le r\}\right) = \Pr\left(|Y_{j+m} - Y_{i+m}| \le r\right),
$$

given the mutual independence of Y_j. It should be noted that this conclusion does not require the condition of identical distribution, as long as the condition of independence is sufficient.

For the simplicity of description, we re-denote Y_{j+m} and Y_{i+m} by two general normally distributed but independent random variables ξ and η whose means are 0 and SDs are 1. The joint probability density functions (PDF) is

$$p(\xi, \eta) = \frac{1}{2\pi\sigma_y^2} e^{-\frac{(\xi-\mu_y)^2+(\eta-\mu_y)^2}{2\sigma_y^2}}$$

and probability is

$$\Pr(|\xi - \eta| \le r) = \iint\limits_{|\xi-\eta|\le r} \frac{1}{2\pi\sigma_y^2} e^{-\frac{(\xi-\mu_y)^2+(\eta-\mu_y)^2}{2\sigma_y^2}} \, d\xi \, d\eta$$

$$= \int_{-\infty}^{\infty} \int_{\eta-r}^{\eta+r} \frac{1}{2\pi\sigma_y^2} e^{-\frac{(\xi-\mu_y)^2+(\eta-\mu_y)^2}{2\sigma_y^2}} \, d\xi \, d\eta.$$

We can then get

$$SampEn(Y; m, r) = -\log \Pr(|Y_{j+1} - Y_{i+1}| \le r)$$
$$= -\log \Pr(|\xi - \eta| \le r)$$
$$= -\log \left(\frac{1}{2\pi\sigma_y^2} \int_{-\infty}^{\infty} \int_{\eta-r}^{\eta+r} e^{-\frac{(\xi-\mu_y)^2+(\eta-\mu_y)^2}{2\sigma_y^2}} \, d\xi \, d\eta \right)$$
$$\overset{\substack{t=\frac{\eta-\mu_y}{\sigma_y}\\s=\frac{\xi-\mu_y}{\sigma_y}}}{=\!=\!=\!=} -\log \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{t-\frac{r}{\sigma_y}}^{t+\frac{r}{\sigma_y}} e^{-\frac{s^2+t^2}{2}} \, ds \, dt \right).$$

Similar to Shannon entropy calculating, after normalize the random variable to have a $\mu = 0$ and $\sigma = 1$, the scale-dependent SD derived for coarse grained signal is $SD(Y^{(\tau)}) = 1/\sqrt{\tau}$. We can get

$$SampEn(Y^{(\tau)}; m, r) = -\log \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{t-r\sqrt{\tau}}^{t+r\sqrt{\tau}} e^{-\frac{s^2+t^2}{2}} \, ds \, dt \right).$$

Since the interval $[t - r\sqrt{\tau}, t + r\sqrt{\tau}]$ increases with τ, the above integral monotonically increases with τ. Accordingly, the negative logarithm based sample entropy $SampEn(Y^{(\tau)}; m, r)$ will monotonically decreases with τ. This is consistent with the aforementioned Shannon entropy-based MSE bias description.

The systematic bias in MSE can be corrected by using a scale adaptive matching threshold. One approach to adjust the threshold is to use $SD^{(\tau)} = SD(0)/\sqrt{\tau}$ for scale τ during $SampEn(Y^{(\tau)}; m, r)$ calculation. This works well for Gaussian signal but may not be effective for other signals if they have extra scale-dependent SD behavior in addition to that induced by the subsampling scale. Finding the theoretical scale-dependent SD equation may not be trivial too. Instead, SD can be directly calculated from the data after each coarse graining. This approach has been proposed in [10].

To demonstrate the systematic bias of MSE and the effeteness of the correction method, we used three synthetic time series with known entropy difference: the Gaussian noise, a $1/f$ noise, and a random walk. The length of time series was $N = 2 \times 10^4$. MSE with and without bias correction were performed.

4. Results

Figure 1 shows the results of MSE with and without bias correction for the three time series (Figure 1a). Parameters used for SampEn calculation were $m = 2$, and $r = 0.15 \times SD$. Without bias correction, MSE produced a monotonically decaying SampEn for Gaussian noise when scale increases. By contrast, SampEn of Gaussian noise stays the same level at different scales after bias correction. The SD bias showed minor effects on SampEn calculation for both 1/f noise and the random walk. Correcting the bias did not dramatically change the SampEn at different scales.
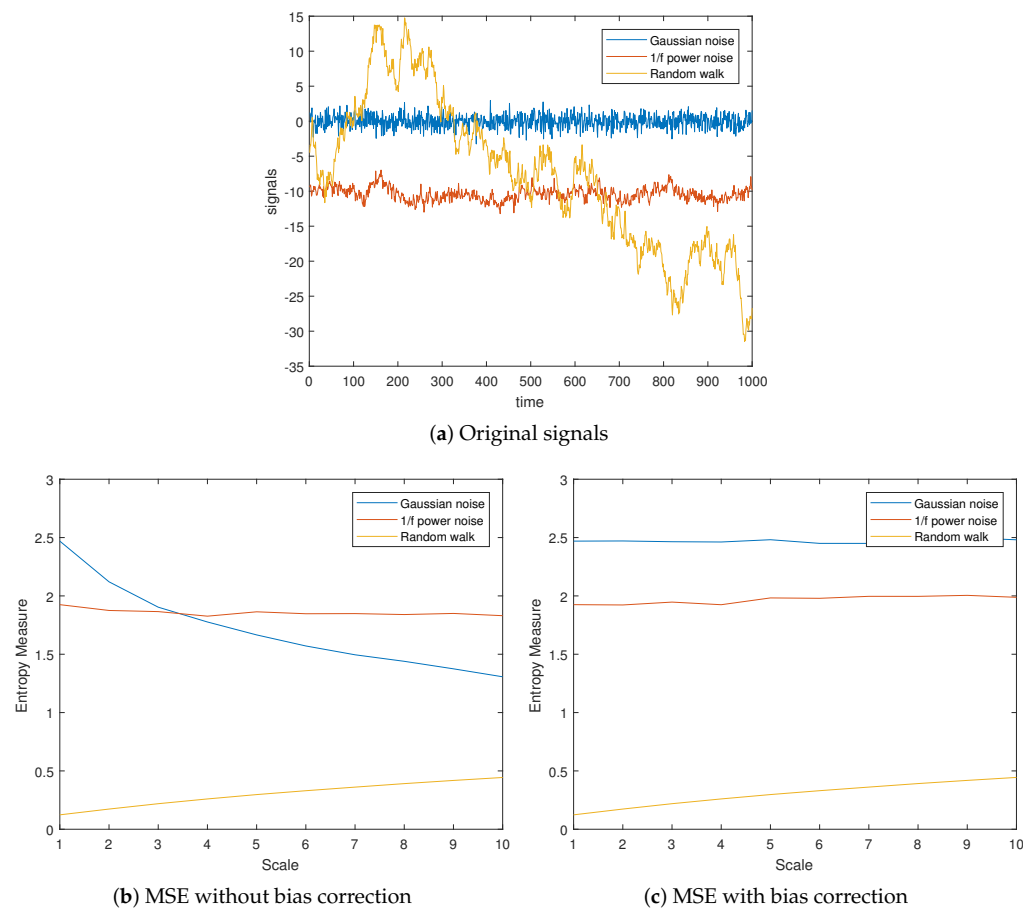

(**a**) Original signals

(**b**) MSE without bias correction

(**c**) MSE with bias correction

Figure 1. Different signals and their SampEn calculated at different scales. (**a**) The original signals before coarse graining; (**b**) MSE without bias correction; (**c**) MSE with bias correction.

5. Discussion and Conclusions

We provided a full mathematical derivation for the systematic bias in MSE introduced by the coarse graining process. We then used synthetic data to show the bias and the correction of it using dynamic SD calculation. Bias correction for Gaussian data MSE calculation works exactly as described by the theoretical descriptions given in this paper. The systematic bias does not appear to be a big issue for the temporally correlated process such as the 1/f noise and random walk. This is because variance of a temporally correlated process does change with the subsampling process if the sampling rate is still higher than the maximal frequency. According to [23], both 1/f noise and random work can be considered special cases of the autoregressive integrated moving average (ARIMA) model. As we derived in Appendix A, an ARIMA model is still an ARIMA model after coarse graining given the condition of that the residuals at different time points are independently and identically distributed (i.i.d.) Gaussian noise. In other words, the moving averaging process will not change the signal variance and will not change SampEn.

While we only showed the results based on one particular set SampEn calcualtion parameters, we included additional figures in Appendix B showing that the bias and the bias correction are still true for other parameters. We did not show the effects of bias correction on real data, but the results shown in the synthetic data should be generalizable to real applications since both the math derivations and the correction process are independent of any specific data but rather general to any dynamic system.

Author Contributions: conceptualization: Z.W.; mathematical derivations: J.L. and Z.W.; experiments: J.L. and Z.W.; draft preparation, J.L.; manuscript review and editing, Z.W. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MSE	multi-scale sample entropy
SampEn	sample entropy
ApEn	Approximate entropy
ARIMA	autoregressive integrated moving average

Appendix A. The Coarse-Grained ARIMA Process

Assume that a time series $\{X_i\}$ can be modeled by an ARIMA process: ARIMA(p,d,q):

$$\left(1 - \sum_{h=1}^{p} \varphi_h L^h\right)(1-L)^d X_i = \left(1 + \sum_{h=1}^{q} \theta_h L^h\right)\varepsilon_i,$$

where $\{\varepsilon_i\}$ are i.i.d. Gaussian noise, L is the lag operator.

Denote the consecutive coarse-grained time series of $\{X_i\}$ by $\{Y_j^{(\tau)}\}$:

$$Y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} X_i.$$

where τ is scale.

Let

$$\epsilon_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} \varepsilon_i,$$

then $\{\epsilon_j^{(\tau)}\}$ are also i.i.d. Gaussian and we can have

$$
\left(1 - \sum_{h=1}^{p} \varphi_h L^h\right)(1-L)^d \left(\frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} X_i\right)
$$

$$
= \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} \left[\left(1 - \sum_{h=1}^{p} \varphi_h L^h\right)(1-L)^d X_i\right]
$$

$$
= \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} \left[\left(1 + \sum_{h=1}^{q} \theta_h L^h\right)\varepsilon_i\right]
$$

$$
= \left(1 + \sum_{h=1}^{q} \theta_h L^h\right)\left(\frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} \varepsilon_i\right).
$$

and

$$
\left(1 - \sum_{h=1}^{p} \varphi_h L^h\right)(1-L)^d Y_j^{(\tau)} = \left(1 + \sum_{h=1}^{q} \theta_h L^h\right)\epsilon_j^{(\tau)}.
$$

This proves that for any scale τ, $\{Y_j^{(\tau)}\}$ is also an ARIMA(p,d,q) process.

Appendix B. Numerical Results on Different m and r

The following Figures A1–A8 show additional MSE calculation results for different SampEn parameters m and r with and without bias correction. N means the length of the time series at scale 1. Theses figures confirmed the systematic bias in the original MSE algorithm for different m and r and the data adaptive correction successfully removed the bias for all assessed signals.
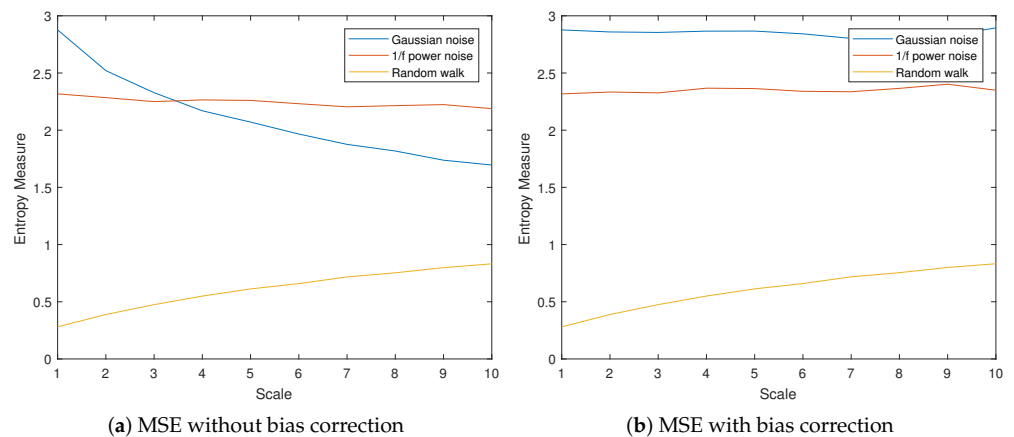

(**a**) MSE without bias correction   (**b**) MSE with bias correction

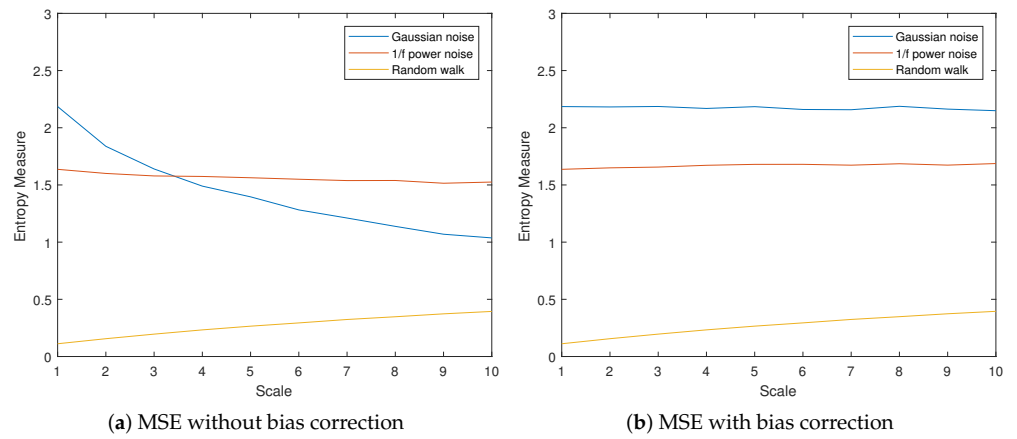**Figure A1.** MSE calculation results with $m = 2$ and $r = 0.1 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction

Figure A2. MSE calculation results with $m = 2$ and $r = 0.2 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction
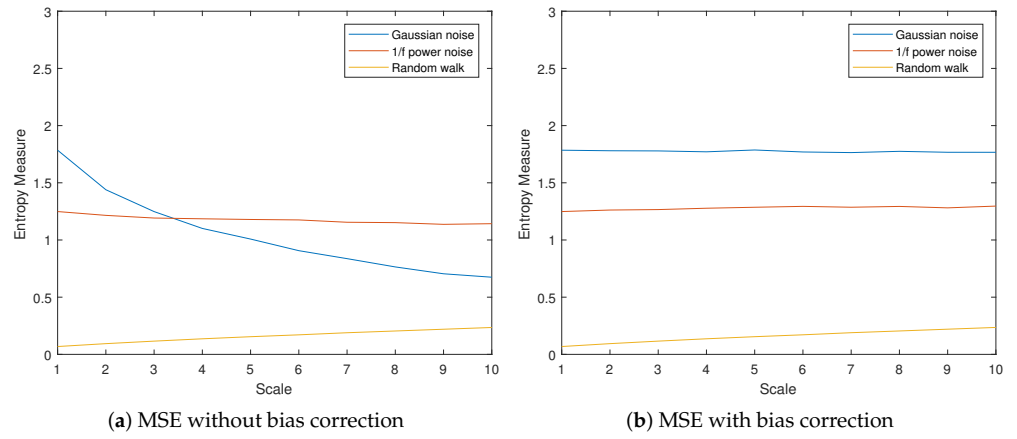
Figure A3. MSE calculation results with $m = 2$ and $r = 0.3 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction
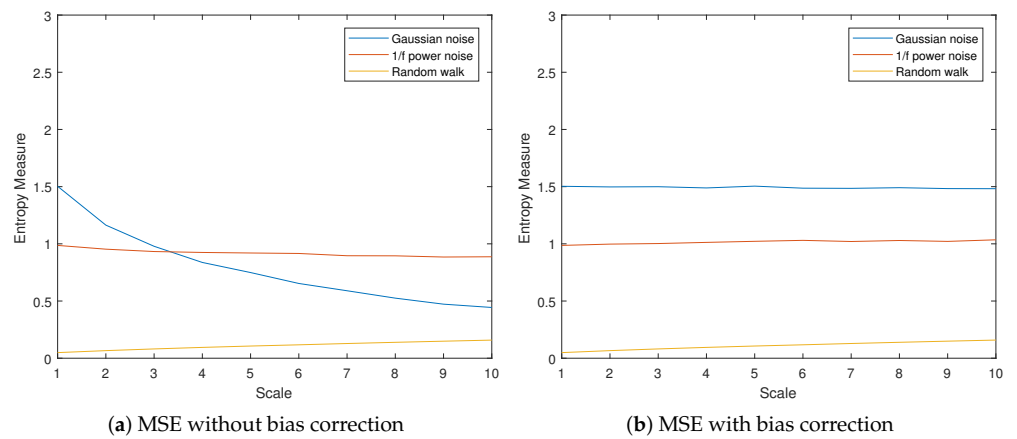
Figure A4. MSE calculation results with $m = 2$ and $r = 0.4 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction
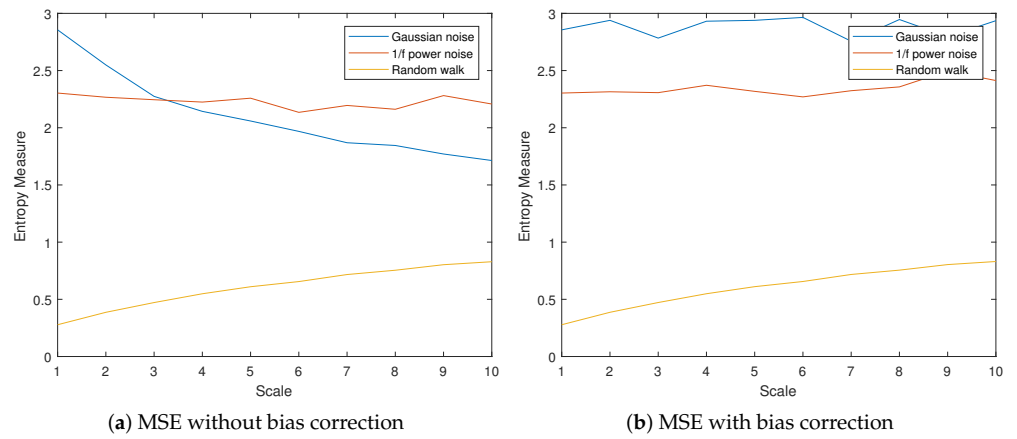
Figure A5. MSE calculation results with $m = 3$ and $r = 0.1 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction
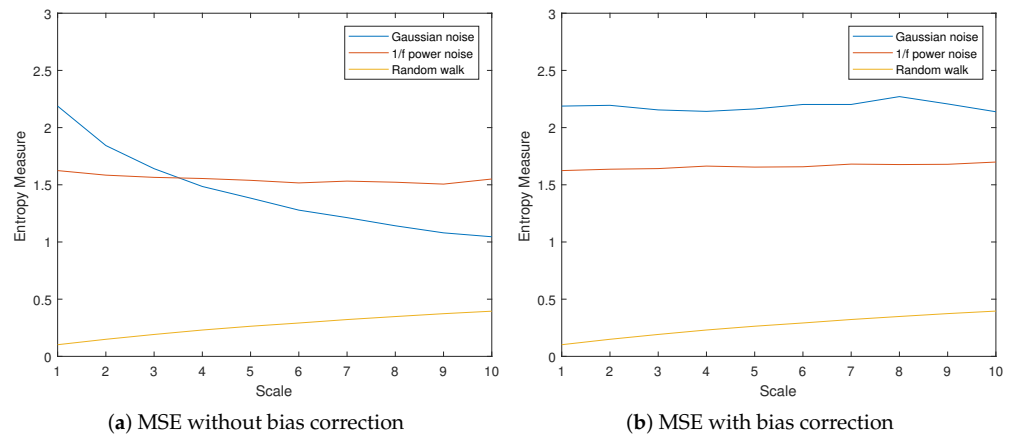
Figure A6. MSE calculation results with $m = 3$ and $r = 0.2 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction

(**b**) MSE with bias correction
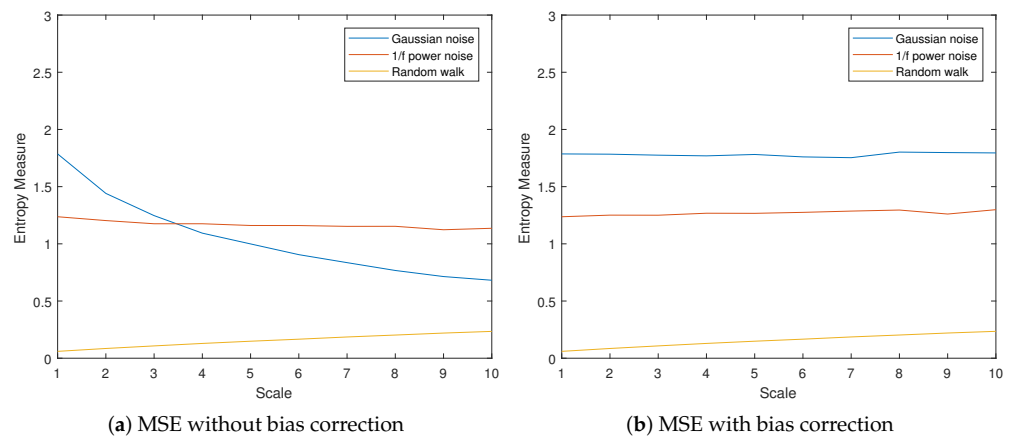
Figure A7. MSE calculation results with $m = 3$ and $r = 0.3 \times SD$. Data length $N = 2 \times 10^4$.

(**a**) MSE without bias correction
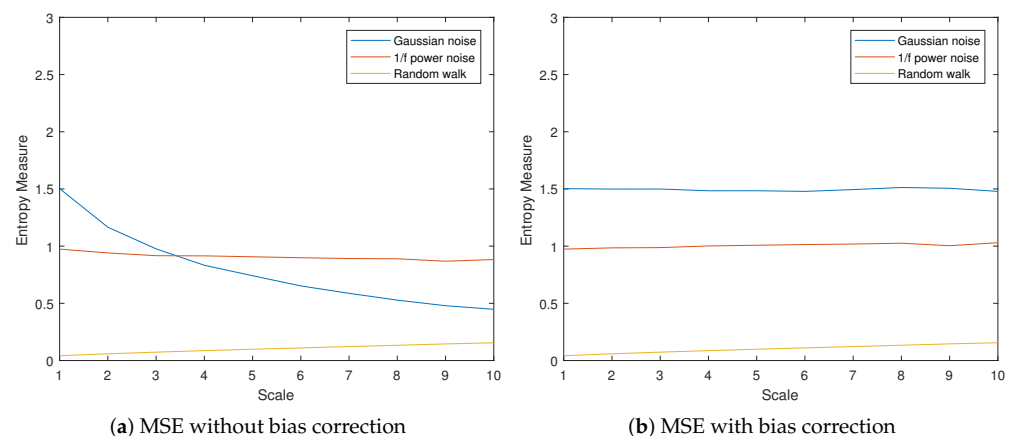
(**b**) MSE with bias correction

Figure A8. MSE calculation results with $m = 3$ and $r = 0.4 \times SD$. Data length $N = 2 \times 10^4$.

References

1. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [CrossRef]
2. Lake, D.E.; Richman, J.S.; Griffin, M.P.; Moorman, J.R. Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2002**, *283*, R789–R797. [CrossRef] [PubMed]
3. Delgado-Bonal, A.; Marshak, A. Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy* **2019**, *21*, 541. [CrossRef]
4. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [CrossRef]
5. Alcaraz, R.; Rieta, J.J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardio-grams. *Biomed. Signal Process. Control* **2010**, *5*, 1–14. [CrossRef]
6. Wang, Z.; Li, Y.; Childress, A.R.; Detre, J.A. Brain entropy mapping using fMRI. *PLoS ONE* **2014**, *9*, e89948. [CrossRef] [PubMed]
7. Sokunbi, M.O. Sample entropy reveals high discriminative power between young and elderly adults in short fMRI data sets. *Front. Neuroinformatics* **2014**, *8*, 69. [CrossRef]
8. Richman, J.S.; Lake, D.E.; Moorman, J.R. Sample entropy. *Methods Enzymol.* **2004**, *384*, 172–184.
9. Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **2015**, *17*, 3110–3123. [CrossRef]
10. Nikulin, V.V.; Brismar, T. Comment on "Multiscale entropy analysis of complex physiologic time series". *Phys. Rev. Lett.* **2004**, *92*, 089803. [CrossRef]
11. Nyquist, H. Certain factors affecting telegraph speed. *Trans. Am. Inst. Electr. Eng.* **1924**, *43*, 412–422. [CrossRef]
12. Hartley, R.V. Transmission of information 1. *Bell Syst. Tech. J.* **1928**, *7*, 535–563. [CrossRef]
13. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
14. Sinai, Y.G. On the notion of entropy of a dynamical system. *Dokl. Russ. Acad. Sci.* **1959**, *124*, 768–771.
15. Grassberger, P.; Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593. [CrossRef]
16. Takens, F. Invariants related to dimension and entropy. *Atas Do* **1983**, *13*, 353–359.
17. Eckmann, J.; Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **1985**, *57*, 617–656. [CrossRef]
18. Pincus, S.M.; Gladstone, I.M.; Ehrenkranz, R.A. A regularity statistic for medical data analysis. *J. Clin. Monit.* **1991**, *7*, 335–345. [CrossRef]
19. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [CrossRef]
20. Zhang, Y.C. Complexity and 1/f noise. A phase space approach. *J. De Phys. I* **1991**, *1*, 971–977. [CrossRef]
21. Fogedby, H.C. On the phase space approach to complexity. *J. Stat. Phys.* **1992**, *69*, 411–425. [CrossRef]
22. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906. [CrossRef] [PubMed]
23. Stadnitski, T. Measuring fractality. *Front. Physiol.* **2012**, *3*, 127. [CrossRef] [PubMed]