

Article

# On the Statistical Properties of Multiscale Permutation Entropy: Characterization of the Estimator's Variance

Antonio Dávalos \*, Meryem Jabloun, Philippe Ravier  and Olivier Buttelli 

Laboratoire Pluridisciplinaire de Recherche en Ingénierie des Systèmes, Mécanique, Énergétique (PRISME), University of Orléans, 45100 Orléans, INSA-CVL, France; meryem.jabloun@univ-orleans.fr (M.J.); philippe.ravier@univ-orleans.fr (P.R.); olivier.buttelli@univ-orleans.fr (O.B.)

\* Correspondence: antonio.davalos-trevino@etu.univ-orleans.fr

Received: 30 March 2019; Accepted: 26 April 2019; Published: 30 April 2019



**Abstract:** Permutation Entropy (PE) and Multiscale Permutation Entropy (MPE) have been extensively used in the analysis of time series searching for regularities. Although PE has been explored and characterized, there is still a lack of theoretical background regarding MPE. Therefore, we expand the available MPE theory by developing an explicit expression for the estimator's variance as a function of time scale and ordinal pattern distribution. We derived the MPE Cramér–Rao Lower Bound (CRLB) to test the efficiency of our theoretical result. We also tested our formulation against MPE variance measurements from simulated surrogate signals. We found the MPE variance symmetric around the point of equally probable patterns, showing clear maxima and minima. This implies that the MPE variance is directly linked to the MPE measurement itself, and there is a region where the variance is maximum. This effect arises directly from the pattern distribution, and it is unrelated to the time scale or the signal length. The MPE variance also increases linearly with time scale, except when the MPE measurement is close to its maximum, where the variance presents quadratic growth. The expression approaches the CRLB asymptotically, with fast convergence. The theoretical variance is close to the results from simulations, and appears consistently below the actual measurements. By knowing the MPE variance, it is possible to have a clear precision criterion for statistical comparison in real-life applications.

**Keywords:** Multiscale Permutation Entropy; ordinal patterns; estimator variance; Cramér–Rao Lower Bound; finite-length signals

## 1. Introduction

Information entropy, originally defined by Shannon [1], has been used as a measure of information content in the field of communications. Several other applications of entropy measurements have been proposed, such as the analysis of physiological electrical signals [2], where a reduction in entropy has been linked to aging [3] and various motor diseases [4]. Another application is the characterization of electrical load behavior, which can be used to perform non-intrusive load disaggregation and to design smart grid applications [5].

Multiple types of entropy measures [6–8] have been proposed in recent years to assess the information content of time series. One notable approach is the Permutation Entropy (PE) [9], used to measure the recurrence of ordinal patterns within a discrete signal. PE is fast to compute and robust even in the presence of outliers [9]. To better measure the information content at different time scales, Multiscale Permutation Entropy (MPE) [10] was formulated as an extension of PE, by using the multiscale approach proposed in [11]. Multiscaling is particularly useful in measuring the information

content in long range trends. The main disadvantage of PE is the necessity of a large data set for it to be reliable [12]. This is especially important in MPE, where the signal length is reduced geometrically at each time scale. Signal-length limitations have been addressed and improved with Composite MPE and Refined Composite MPE [13].

PE theory and properties have been extensively explored [14–16]. However, there is still a lack of understanding regarding the statistical properties of MPE. In our previous work [17], we already derived the expected value of the MPE measurement, taking into consideration the time scale and the finite-length constraints. We found the MPE expected value to be biased. Nonetheless, this bias depends only on the MPE parameters, and it is constant with respect to the pattern probability distribution. In practice, the MPE bias does not depend on the signal [17], and thus, can be easily corrected.

In the present article, our goal is to continue this statistical analysis by obtaining the variance of the MPE estimator by means of Taylor series expansion. We develop an explicit equation for the MPE variance. We also obtain the Cramér–Rao Lower Bound (CRLB) of MPE, and compare it to our obtained expression to assess its theoretical efficiency. Lastly, we compare these results with simulated data with known parameters. This gives a better understanding of the MPE statistic, which is helpful in the interpretation of this measurement in real-life applications. By knowing the precision of the MPE, it is possible to take informed decisions regarding experimental design, sampling, and hypothesis testing.

The remainder of the article is organized as follows: In Section 2, we lay the necessary background of PE and MPE, the main derivation of the MPE variance, and the CRLB. We also develop the statistical model to generate the surrogate data simulations for later testing. In Section 3, we show and discuss the results obtained, including the properties of the MPE variance, its theoretical efficiency, and similarities with our simulations. Finally, in Section 4, we add some general remarks regarding the results obtained.

## 2. Materials and Methods

In this section, we establish the concepts and tools necessary for the derivation of the MPE model. In Section 2.1 we review the formulation of PE and MPE in detail. In Section 2.2, we show the derivation of the MPE variance. In Section 2.3, we derive the expression for the CRLB of the MPE, and we compare it to the variance. Finally, in Section 2.4, we explain the statistical model used to generate surrogate signals, which are used to test the MPE model.

### 2.1. Theoretical Background

#### 2.1.1. Permutation Entropy

PE [9] measures the information content by counting the ordinal patterns present within a signal. An ordinal pattern is defined as the comparison between the values of adjacent data points in a segment of size  $d$ , known as the embedded dimension. For example, for a discrete signal of length  $N$ ,  $x_1, \dots, x_n, \dots, x_N$ , and  $d = 2$ , only two possible patterns can be found within the series,  $x_n < x_{n+1}$  and  $x_n > x_{n+1}$ . For  $d = 3$ , there are six possible patterns present, as shown in Figure 1. In general, for any integer value of  $d$ , there exists  $d$  factorial ( $d!$ ) possible patterns within any given signal segment. To calculate PE, we must first obtain the sample probabilities within the signal, by counting the number of times a certain pattern  $i = 1, \dots, d!$  occurs. This is formally expressed as follows:

$$P(\pi_i) = \frac{\#\{n | n \leq N - (d - 1), (x_n, \dots, x_{n+d-1}) \text{ type } \pi_i\}}{N - (d - 1)} = \hat{p}_i. \quad (1)$$

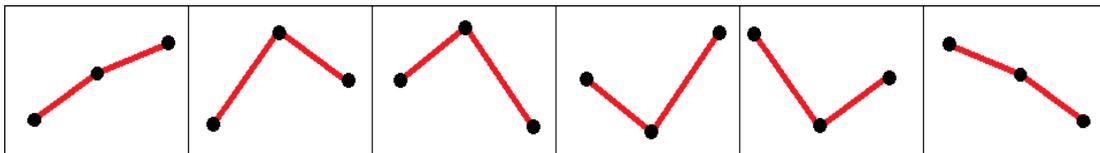
where  $\pi_i$  is the label of a particular ordinal pattern  $i$ , and  $\hat{p}_i$  is the estimated pattern probability. Some authors [15] also introduce a downsampling parameter in Equation (1), to address the analysis of an oversampled signal. For the purposes of this article, we assume a properly sampled signal, and thus, we do not include this parameter.

Using these estimations, we can apply the entropy definition [1] to obtain the PE of the signal

$$\hat{\mathcal{H}} = \frac{-1}{\ln(d!)} \sum_{i=1}^{d!} \hat{p}_i \ln \hat{p}_i. \quad (2)$$

where  $\hat{\mathcal{H}}$  is the estimated normalized PE from the data.

PE is a very simple and fast estimator to calculate, and it is invariant to non-linear monotonous transformations [15]. It is also convenient to note that we need no prior assumptions on the probability distribution of the patterns, which makes PE a very robust estimator. The major disadvantage in PE involves the length of the signal, where we need  $N \gg d!$  for PE to be meaningful. This imposes a practical constraint in the use of PE for short signals, or in conditions where the data length is reduced.



**Figure 1.** All possible patterns for embedded dimension  $d = 3$ , from a three-point sequence  $\{x_n, x_{n+1}, x_{n+2}\}$ . The patterns represented, from left to right, are  $\pi_1 : x_n < x_{n+1} < x_{n+2}$ ,  $\pi_2 : x_n < x_{n+2} < x_{n+1}$ ,  $\pi_3 : x_{n+2} < x_n < x_{n+1}$ ,  $\pi_4 : x_{n+1} < x_n < x_{n+2}$ ,  $\pi_5 : x_{n+1} < x_{n+2} < x_n$ , and  $\pi_6 : x_{n+2} < x_{n+1} < x_n$ . The difference in amplitude between data points does not affect the pattern, as long as the order is preserved.

### 2.1.2. Multiscale Permutation Entropy

The MPE consists of applying the classical PE analysis on coarse-grained signals [10]. First, we define  $m$ , as the time scale parameter for the MPE analysis. The coarse-graining procedure consists in dividing the original signal into adjacent, non-overlapping segments of size  $m$ , where  $m$  is a positive integer less than  $N$ . The data points within each segment are averaged to produce a single data point, which represents the segment at the given time scale [11].

$$x_k^{(m)} = \frac{1}{m} \sum_{j=m(k-1)+1}^{km} x_j. \quad (3)$$

MPE consists in applying the PE procedure in Equation (2) on the coarse-grained signals in Equation (3) for different time scales  $m$ . This technique allows us to measure the information inside longer trends and time scales, which is not possible using PE. Nonetheless, the resulting coarse-grained signals have a length of  $N/m$ , which limits the analysis. Moreover, for a sufficiently large  $m$ , the condition  $N/m \gg d!$  is eventually not satisfied.

At this point, it is important to discuss some constraints regarding the interaction between time scale  $m$  and the signal length  $N$ . First, strictly speaking,  $N/m$  must be an integer. In practice, the coarse-grained signal length will be the integer number immediately below  $N/m$ . Second, the proper domain of  $m$  is  $(0, N)$ , as the segments size, at most, can be the same length as the signal itself. It is handy to use a normalized scale  $\frac{m}{N}$  with domain  $(0, 1)$  which does not change between signals. This normalized scale is the inverse of the coarse-grained signal length. Taking the length constraints from the previous section, this means that  $\frac{m}{N} \ll \frac{1}{d!}$ . For  $d = 2$ , a normalized scale of  $\frac{m}{N} = \frac{1}{2}$  will result in a coarse-grained signal of 2 data points, which is not meaningful for this analysis. For  $d = 3$ , the normalized scale must be significantly less than  $\frac{1}{6}$ , which corresponds to a coarse-grained signal of six elements. Therefore, for practical applications, we restrict our analysis for values of  $\frac{m}{N}$  close to zero, by selecting a small time scale  $m$ , a large signal length  $N$ , or both.

### 2.2. Variance of MPE Statistic

The calculated MPE can be interpreted as a sample statistic that estimates the true entropy value, with an associated expected value, variance, and bias, for each time scale. We have previously developed the calculation of the expected value of the MPE in [17], where the bias has been found to be independent from the pattern probabilities. We expand on this result by first proposing an explicit equation to the MPE statistic, and then we formulate the variance of MPE, as a function of the true pattern probabilities and time scale.

For the following development, we use  $H$ , the non-normalized version of Equation (2), for convenience.

$$\hat{H} = \hat{\mathcal{H}} \ln(d!) = - \sum_{i=1}^{d!} \hat{p}_i \ln \hat{p}_i. \tag{4}$$

By taking the Taylor expansion of Equation (4) on a coarse-grained signal in Equation (3), we get

$$\hat{H} = H - \frac{m}{N} \sum_{i=1}^{d!} (1 + \ln(p_i)) \Delta Y_i - \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k(k+1)} \left(\frac{m}{N}\right)^{k+1} \sum_{i=1}^{d!} \frac{\Delta Y_i^{k+1}}{p_i^k} \tag{5}$$

where  $\hat{H}$  is the MPE estimator,  $H$  is the true unknown MPE value,  $m$  is the time scale,  $N$  is the signal length, and  $d$  the embedded dimension. The probabilities  $p_i$  correspond to the true probabilities (unknown parameters) of each pattern, and  $\Delta Y_i$  correspond to the random part in the multinomially (Mu) distributed frequency of each pattern,

$$\begin{aligned} Y_i &= \frac{N}{m} \hat{p}_i = \frac{N}{m} p_i + \Delta Y_i, \\ \{Y_1, \dots, Y_{d!}\} &\sim \text{Mu} \left( \frac{N}{m}, p_1, \dots, p_{d!} \right) \end{aligned} \tag{6}$$

$Y_i$  being the number of pattern counts of type  $i$  in the signal.

If we take into consideration the length constraints discussed in Section 2.1.2, we know that the normalized time scale  $\frac{m}{N}$  is very close to zero. This implies that the high-order terms of Equation (5) will quickly vanish for increasing values of  $k$ . Therefore, we propose to make the simplest approximation of Equation (5) by taking only the term  $k = 1$ . By doing this, we get the following expression:

$$\hat{H} \approx H - \frac{m}{N} \sum_{i=1}^{d!} (1 + \ln(p_i)) \Delta Y_i - \frac{1}{2} \left(\frac{m}{N}\right)^2 \sum_{i=1}^{d!} \frac{\Delta Y_i^2}{p_i}. \tag{7}$$

Using our previous results involving MPE [17], we know the expected value of Equation (7) is

$$E[\hat{H}] \approx H - \frac{d! - 1}{2} \left(\frac{m}{N}\right). \tag{8}$$

The statistic presents a bias that does not depend on the pattern probabilities  $p_i$ , and thus, can be easily corrected for any signal.

Now, we move further and calculate the variance of the MPE estimator. First, it is convenient to express Equation (4) in vectorial form.

$$\hat{H} = -\hat{\mathbf{l}}^T \hat{\mathbf{p}} \tag{9}$$

where

$$\begin{aligned} \hat{\mathbf{p}} &= [\hat{p}_1, \dots, \hat{p}_{d!}]^T \\ \hat{\mathbf{l}} &= [\ln(\hat{p}_1), \dots, \ln(\hat{p}_{d!})]^T. \end{aligned} \tag{10}$$

$\hat{\mathbf{p}}$  being the column vector of pattern probability estimators,  $\hat{\mathbf{l}}$  the column vector of the logarithm of each pattern probability estimator, and  $T$  is the transpose symbol. We can now rewrite Equation (7) as

$$\hat{H} \approx H - \frac{m}{N} (\mathbf{1} + \mathbf{l})^T \Delta \mathbf{Y} - \frac{1}{2} \left(\frac{m}{N}\right)^2 (\mathbf{p}^{\circ-1})^T \Delta \mathbf{Y}^{\circ 2} \tag{11}$$

where

$$\begin{aligned} \mathbf{p} &= [p_1, \dots, p_{d!}]^T & \mathbf{p}^{\circ-1} &= [p_1^{-1}, \dots, p_{d!}^{-1}]^T \\ \mathbf{1} &= [1, \dots, 1]^T & \mathbf{l} &= [\ln(p_1), \dots, \ln(p_{d!})]^T \\ \mathbf{p}^{\circ 2} &= [p_1^2, \dots, p_{d!}^2]^T & \Delta \mathbf{Y}^{\circ 2} &= [\Delta Y_1^2, \dots, \Delta Y_{p_{d!}}^2]^T. \end{aligned} \tag{12}$$

The circle notation  $\circ$  represents the Hadamard power (element-wise). Now, we can obtain the variance of the MPE estimator (11),

$$\begin{aligned} var(\hat{H}) &= E[\hat{H}^2] - (E[\hat{H}])^2 \\ &\approx H^2 + \left(\frac{m}{N}\right)^2 (\mathbf{1} + \mathbf{l})^T E[\Delta \mathbf{Y} \Delta \mathbf{Y}^T] (\mathbf{1} + \mathbf{l}) - \left(\frac{m}{N}\right)^2 (\mathbf{p}^{\circ-1})^T E[\Delta \mathbf{Y}^{\circ 2}] H \\ &\quad + \left(\frac{m}{N}\right)^3 (\mathbf{1} + \mathbf{l})^T E[\Delta \mathbf{Y} (\Delta \mathbf{Y}^{\circ 2})^T] (\mathbf{p}^{\circ-1}) + \frac{1}{4} \left(\frac{m}{N}\right)^4 (\mathbf{p}^{\circ-1})^T E[\Delta \mathbf{Y}^{\circ 2} (\Delta \mathbf{Y}^{\circ 2})^T] (\mathbf{p}^{\circ-1}) \\ &\quad + \left(\frac{m}{N}\right) (d! - 1) H - \frac{1}{4} \left(\frac{m}{N}\right)^2 (d! - 1)^2 - H^2. \end{aligned} \tag{13}$$

Now, we know that  $E[\Delta \mathbf{Y} \Delta \mathbf{Y}^T]$  is the Covariance matrix of  $\Delta \mathbf{Y}$ , which is the multinomial random variable defined in Equation (6). The matrix  $E[\Delta \mathbf{Y} (\Delta \mathbf{Y}^{\circ 2})^T]$  is the Coskewness matrix, and  $E[\Delta \mathbf{Y}^{\circ 2} (\Delta \mathbf{Y}^{\circ 2})^T]$  is the Cokurtosis. We know that (see Appendix A),

$$E[\Delta \mathbf{Y} \Delta \mathbf{Y}^T] = \frac{N}{m} (diag(\mathbf{p}) - \mathbf{p} \mathbf{p}^T) \tag{14}$$

$$E[\Delta \mathbf{Y} (\Delta \mathbf{Y}^{\circ 2})^T] = 2 \frac{N}{m} (\mathbf{p}^{\circ 2} \mathbf{p}^T - diag(\mathbf{p}^{\circ 2})) + \frac{N}{m} (diag(\mathbf{p}) - \mathbf{p} \mathbf{p}^T) \tag{15}$$

$$\begin{aligned} E[\Delta \mathbf{Y}^{\circ 2} (\Delta \mathbf{Y}^{\circ 2})^T] &= 3 \frac{N}{m} \left(\frac{N}{m} - 2\right) \mathbf{p}^{\circ 2} (\mathbf{p}^{\circ 2})^T - \frac{N}{m} \left(\frac{N}{m} - 2\right) (\mathbf{p}^{\circ 2} \mathbf{p}^T + \mathbf{p} (\mathbf{p}^{\circ 2})^T) \\ &\quad + \left(\frac{N}{m}\right)^2 \mathbf{p} \mathbf{p}^T - 4 \frac{N}{m} \left(\frac{N}{m} - 2\right) diag(\mathbf{p}^{\circ 3}) + 2 \frac{N}{m} \left(\frac{N}{m} - 3\right) diag(\mathbf{p}^{\circ 2}) \\ &\quad + \frac{N}{m} (diag(\mathbf{p}) - \mathbf{p} \mathbf{p}^T) \end{aligned} \tag{16}$$

where  $diag(\mathbf{p})$  is a diagonal matrix, where the diagonal elements are the elements of  $\mathbf{p}$ .

We can further summarize the covariance matrix (14) as follows:

$$\frac{N}{m} (diag(\mathbf{p}) - \mathbf{p} \mathbf{p}^T) = \frac{N}{m} \Sigma_{\mathbf{p}}. \tag{17}$$

We also rewrite the following term in Equation (13),

$$(\mathbf{p}^{\circ-1})^T E[\Delta \mathbf{Y}^{\circ 2}] = \left(\frac{N}{m}\right) (\mathbf{p}^{\circ-1})^T (\mathbf{p} - \mathbf{p}^{\circ 2}) = \left(\frac{N}{m}\right) (d! - 1). \tag{18}$$

By combining Equations (14) to (18) explicitly in Equation (13), we get the expression,

$$var(\hat{H}) \approx \left(\frac{m}{N}\right) \mathbf{l}^T \Sigma_{\mathbf{p}} \mathbf{l} + \left(\frac{m}{N}\right)^2 \left(\mathbf{1}^T \mathbf{l} + d! H + \frac{1}{2} (d! - 1)\right) + \frac{1}{4} \left(\frac{m}{N}\right)^3 \left(\mathbf{1}^T \mathbf{p}^{\circ-1} - (d!^2 + 2d! - 2)\right). \tag{19}$$

We note that Equation (19) is a cubic polynomial respect to the normalized scale  $m/N$ . Since we are still working in the region where  $m/N$  is close to (but not including) zero, we propose to further

simplify this expression using only the linear term. This means that we can approximate Equation (19) as follows:

$$var(\hat{H}) \approx \left(\frac{m}{N}\right) \mathbf{l}^T \Sigma_p \mathbf{l} = \frac{m}{N} \left( \sum_{i=1}^{d!} p_i \ln^2(p_i) - \sum_{i=1}^{d!} \sum_{j=1}^{d!} p_i p_j \ln(p_i) \ln(p_j) \right) = \frac{m}{N} \left( \sum_{i=1}^{d!} p_i \ln^2(p_i) - H^2 \right). \quad (20)$$

We note that Equation (20) will be equal to zero for a perfectly uniform pattern distribution (which yields a maximum PE). In this particular case, Equation (20) will not be a good approximation for the MPE variance, and we will need, at least, the quadratic term in Equation (19). Except for extremely high or low values of MPE, we expect the variance linear approximation to be accurate. We discuss more properties of this statistic in the Section 3.2.

### 2.3. MPE Cramér–Rao Lower Bound

In this section, we compare the MPE variance (20) to the CRLB, to test the efficiency of our estimator. The CRLB is defined as

$$CRLB(H) = \frac{[1 - B'(H)]^2}{I(H)} \leq var(\hat{H}) \quad (21)$$

where  $B$  is the bias of the MPE expected value from Equation (8) and

$$B'(H) = \frac{dB}{dH} = -\frac{d}{dH} \left( \frac{d! - 1}{2} \frac{m}{N} \right) = 0$$

$$I(H) = -E \left[ \frac{\partial^2 \ln(f_H(\mathbf{y}; \mathbf{p}))}{\partial H^2} \right] \quad (22)$$

where  $I(H)$  is the Fisher Information, and  $f_H(\mathbf{y}; \mathbf{p})$  is the probability distribution function of  $H$ .

Although we do not have an explicit expression for  $f_H$ , we can express  $CRLB(H)$  as a function of  $CRLB(\mathbf{p})$  as follows [18]:

$$CRLB(H) = \left( \frac{\partial H}{\partial \mathbf{p}} \right)^T CRLB(\mathbf{p}) \frac{\partial H}{\partial \mathbf{p}} \quad (23)$$

where

$$\frac{\partial H}{\partial \mathbf{p}} = \left[ \frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_{d!}} \right]^T \quad (24)$$

$$CRLB(\mathbf{p}) = \mathbf{I}(\mathbf{p})^{-1} \leq cov_p(\hat{\mathbf{p}}). \quad (25)$$

$\mathbf{I}(\mathbf{p})$  is the Fisher information matrix for  $\mathbf{p}$ , which we know has a multinomial distribution related to Equation (6). Each element of  $\mathbf{I}(\mathbf{p})$  is defined as

$$I_{j,k}(\mathbf{p}) = -E \left[ \frac{\partial^2}{\partial p_j \partial p_k} \ln(f_{\hat{\mathbf{p}}}(\hat{\mathbf{p}}; \mathbf{p})) \right]. \quad (26)$$

where  $f_{\hat{\mathbf{p}}}(\hat{\mathbf{p}}; \mathbf{p})$  is the probability distribution of  $\hat{\mathbf{p}}$ , which is identically distributed to Equation (6) (for the full calculation of  $CRLB(H)$ , see Appendix B). Thus, by obtaining the inverse of  $\mathbf{I}(\mathbf{p})$  and all partial derivatives of  $H$  with respect to each element of  $\mathbf{p}$ , we obtain the  $CRLB(H)$ .

$$CRLB(H) = \frac{m}{N} \left( \sum_{i=1}^{d!} p_i \ln^2 p_i - H^2 \right) = \frac{m}{N} \mathbf{l}^T \Sigma_p \mathbf{l}. \quad (27)$$

As we recall from our results in Equations (19) and (20), the CRLB corresponds to the first term of the Taylor series expansion of the MPE variance. The high-order terms in Equation (19) increase

the MPE variance above the CRLB. For small values of  $\frac{m}{N}$ , the higher-order terms in Equations (20) become neglectable, which make the MPE variance approximation converge to  $CRLB(H)$ .

#### 2.4. Simulations

To test the MPE variance, we need an appropriate benchmark. We need to design a proper signal model with the following goals in mind: First, the model must preserve the pattern probabilities across all the signal generated; second, the function must have the pattern probability as an explicit parameter, easily modifiable for testing. The following equation satisfies these criteria for dimension  $d = 2$ :

$$x_n = x_{n-1} + \epsilon_n - \delta(p) \quad (28)$$

where

$$p = \frac{1}{2} \left( 1 - \operatorname{erf} \left( \frac{\delta(p)}{\sqrt{2}} \right) \right) \quad (29)$$

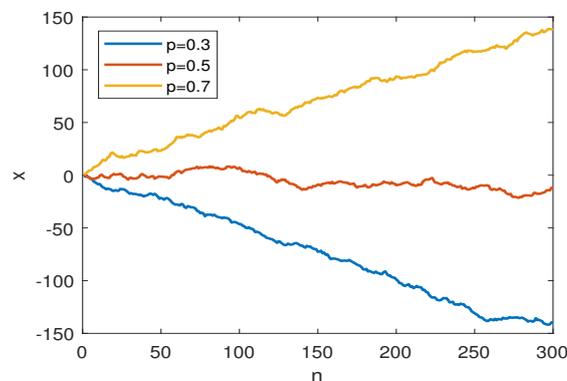
$$\delta(p) = \sqrt{2} \operatorname{erf}^{-1}(1 - 2p). \quad (30)$$

This is a non-stationary process, with a trend function  $\delta(p)$ .  $\epsilon$  is a Gaussian noise term with variance  $\sigma^2 = 1$ , without loss of generality. Although different values of  $\sigma^2$  will indeed modify the trend function, it will not affect the pattern probabilities in the simulation, as PE is invariant to non-linear monotonous transformations [15].

For dimension  $d = 2$ ,  $p = p_1 = P(x_n < x_{n+1})$  and  $1 - p = p_2 = P(x_n > x_{n+1})$ , which are the probabilities of each of the two possible patterns. The formulation of  $\delta(p)$  comes from the Gaussian Complementary Cumulative Distribution function, taking  $p$  as a parameter. This guarantees that we can directly modify the pattern probabilities  $p$  and  $1 - p$  for simulation.

Although  $p$  is not invariant at different time scales, it is consistent within each coarse-grained signal, which suffices for our purposes. We chose to restrict our simulation analysis to the embedded dimension  $d = 2$ . Although our theoretical work holds for any value of  $d$  (see Section 2.2), it is difficult to visualize the results at higher dimensions.

This surrogate model (28) was implemented in the Matlab environment. For the test, we generated 1000 signals each, for 99 different values for  $p = 0.01, 0.02, \dots, 0.99$ . The signal length was set to  $N = 1000$ . Some sample paths are shown in Figure 2. These signals were then subject to the coarse-graining procedure (3) for time scale  $m = 1, \dots, 10$ . The MPE measurement was taken for each coarse-grained simulated signal using Equation (2). Finally, we obtained the mean and variance of the resulting MPE's for each scale. This simulation results are discussed in Section 3.2.



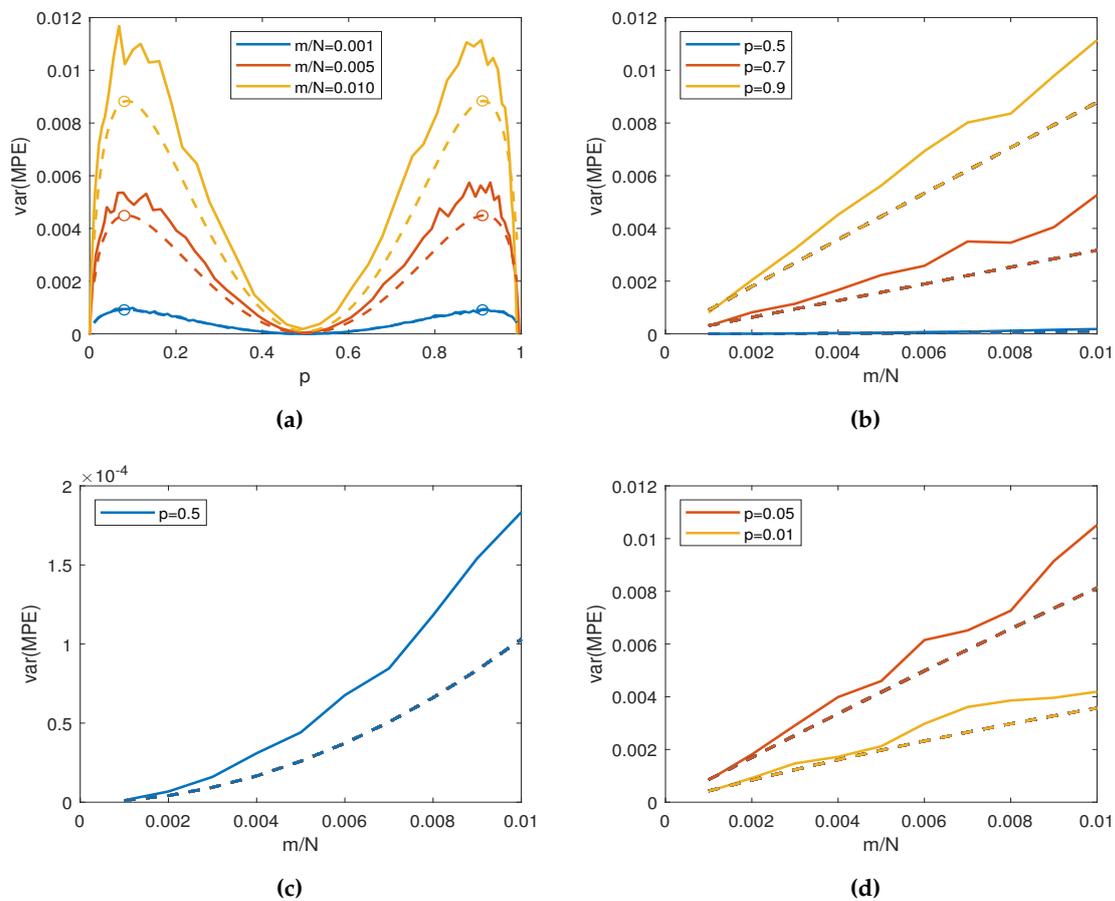
**Figure 2.** Simulated paths for MPE testing, where  $p$  is the probability of  $x_n < x_{n+1}$  for dimension  $d = 2$ . The graph shows sample paths for  $p = 0.3$ ,  $p = 0.5$ ,  $p = 0.7$

### 3. Results and Discussion

In this section, we explore the results from the theoretical MPE variance. In Section 3.1, we contrast the results from the model against the MPE variance measured from the simulations. In Section 3.2, we discuss these findings in detail.

#### 3.1. Results

Here, we compare the theoretical results with the surrogate data obtained by means of the procedure described in Section 2.4. We use the cubic model from Equation (19) instead of the linear approximation in Equation (20), to take in consideration non-linear effects that could arise from simulations. Figure 3 shows the theoretical predictions in dotted lines, and simulation measurements in solid lines.



**Figure 3.**  $var(\hat{H})$  for embedded dimension  $d = 2$  from theory (dotted lines) and simulations (solid lines). (a)  $var(\hat{H})$  vs. pattern probability  $p$  for different normalized scales  $m/N$ . (b)  $var(\hat{H})$  vs.  $m/N$  for different values of  $p$ . (c)  $var(\hat{H})$  vs.  $m/N$  at  $p = 0.5$ , which corresponds to maximum entropy. (d)  $var(\hat{H})$  vs.  $m/N$  at with small  $p$ , which approaches minimum entropy.

Figure 3a shows the variance of the MPE (19) as a function of  $p$  for  $d = 2$ . The lines correspond to a normalized time scale of  $m/N = 0.001, 0.005$ , and  $0.010$ . The variance presents symmetry with respect to the middle value of  $p = 0.5$ . This is to be expected, as for  $d = 2$ , the variance has only one degree of freedom. The structure is preserved, albeit scaled, for different values of  $m$ .

Figure 3b–d show the MPE variance versus the normalized time scale, for different values of pattern probability  $p$ . As we can see in Figure 3b, the variance increases linearly with respect to the normalized scale  $m/N$  at the positive vicinity of zero, as predicted in Equation (20). This is not the case for when  $p = 0.5$  (maximum entropy), as shown in Figure 3c, where both the theory and simulations

show a clear non-linear tendency. Finally, Figure 3d shows the case where we have extreme pattern probability distributions, with entropy close to zero. Although we use the complete cubic model (19), the predicted curves are almost linear.

In general, we can observe that the simulation results have greater variance than the prediction of the model. The real values from the simulations correspond to the sample variance from the signals, calculated from  $\hat{p}$  instead of  $p$ . Nonetheless, the simulations and theoretical graphs have the same shape. It is interesting to note that the discrepancies increase with the scale. This effect is addressed in Section 3.2.

### 3.2. Discussion

It is interesting to explore the particular structure of the variance. As we can observe from Figure 3b, the MPE variance increases linearly with respect to the time scale for a wide range of pattern distributions, as described in Equation (20). This is even true with highly uneven distributions, where the entropy is very close to zero, as shown in Figure 3d. Nonetheless, when we observe the expression (20), the equation collapses to zero when all pattern probabilities  $p_i$  are the same. For any embedded dimension  $d$ , all probabilities  $p_i = \frac{1}{d!}$ . This particular pattern probability distribution is the discrete uniform distribution, which yield to the maximum possible entropy in Equation (2). As we can observe from Figure 3c, the linear approximation in Equation (19) is not enough to estimate the variance in this case. Nonetheless, the results suggest a quadratic increase. This agrees with previous results by Little and Kane [16], where they characterized the classical normalized PE variance for white noise under finite-length constraints.

$$var(\hat{\mathcal{H}}) \approx \frac{d! - 1}{2(\ln d!)^2} \frac{1}{N^2} \tag{31}$$

This result matches our model in Equation (19) (taking the quadratic term) for the specific case of uniform pattern distribution and time scale  $m = 1$ . This suggest that, when we approach the maximum entropy, the quadratic approximation is necessary.

Contrary to the bias in the expected value of MPE [17], the variance is strongly dependent on the pattern probabilities present in the signal, as shown in Figure 3a. For the MPE with embedded dimension  $d = 2$ , the variance of MPE has a symmetric shape around equally probable patterns. We observe, for a fixed time scale, that the variance increases the further we deviate from the center, and sharply decreases for extreme probabilities. The variance presents its lowest points at the center and the extremes of the pattern probability distributions, which corresponds to the points where the entropy is maximum and minimum, respectively. For  $d = 2$ , we can calculate the variance (20) more explicitly,

$$var(\hat{H}) \approx \left(\frac{m}{N}\right) \mathbf{I}^T \Sigma_p \mathbf{I} |_{d=2} = \left(\frac{m}{N}\right) p(1-p) \ln^2\left(\frac{p}{1-p}\right). \tag{32}$$

It is evident that the zeros of this equation correspond to  $p = 0$ ,  $p = 1$  (points of minimum entropy), and  $p = 0.5$  (maximum entropy). We can get the critical points by calculating the first derivative of Equation (32) with respect to  $p$

$$\left(\frac{m}{N}\right) \frac{d}{dp} \left(\mathbf{I}^T \Sigma_p \mathbf{I} |_{d=2}\right) = \left(\frac{m}{N}\right) \ln\left(\frac{p}{1-p}\right) \left((1-2p) \ln\left(\frac{p}{1-p}\right) + 2\right) = 0 \tag{33}$$

which is equal to zero to get the extreme points. From the first term of Equation (33), we obtain the critical point  $p = 0.5$ , which is a minimum. If the second term of the equation is equal to zero, we need to solve the transcendental function

$$\ln\left(\frac{p}{1-p}\right) = \frac{2}{2p-1}. \tag{34}$$

Numerically, we found the maximum points to be  $p = 0.083$  and  $p = 0.917$ , as shown in Figure 3a. Both these values yield to a normalized PE value of  $\hat{\mathcal{H}} = 0.413$ . This implies that, when we obtain values of MPE close to this value, we will have a region with maximum variance. This effect arises directly from the pattern probability distribution. Hence, in the region around  $\hat{\mathcal{H}} = 0.413$  for  $d = 2$ , we will have maximum estimation uncertainty, even before factoring the finite length of the signal, or the time scale. Therefore, Equation (20) implies an uneven variance across all possible values of the entropy measurement, regardless of time scale and embedded dimension. It also implies a region where this variance is maximum.

Lastly, as noted in Section 2.3, the MPE variance (20) approaches the CRLB for small values of  $m$ . This is further supported by the simulation variance MPE measurements shown in Figure 3, which are consistently above the theoretical prediction. We can attribute this effect to the number of iterations of the testing model in Equation (28), where an increasing number of repetitions will yield to a more precise MPE estimation, with a reduced variance. Since we already include the effect of the time scale in Equations (19) and (20), the difference between the theoretical results and the simulations does not come directly from the signal length or the pattern distribution.

#### 4. Conclusions

By following on from our previous work [17], we further explored and characterized the MPE statistic. By using a Taylor series expansion, we were able to obtain an explicit expression of the MPE variance. We also derived the Cramér–Rao Lower bound of the MPE, and compared it to our obtained expression. Finally, we proposed a suitable signal model for testing our results against simulations.

By analyzing the properties of the MPE variance graph, we found the estimator to be symmetric around the point of equally probable patterns. Moreover, the estimator is minimum in both the points of maximum and minimum entropy. This implies, first, that the variance of the MPE is dependent on the MPE estimation itself. In regions where the entropy measurement is near the maximum or minimum, the estimation will have a minimum variance. On the other hand, there is an MPE measurement where the variance will have a maximum. This effect comes solely from the pattern distribution, and not from the signal length or the MPE time scale. We should take in account this effect when comparing real entropy measurements, as it could affect the interpretation of statistically significant difference.

Regarding the time scale, the MPE variance linear approximation is sufficiently accurate for almost all pattern distributions, provided that the time scale is small compared to the signal length. An important exception to this is the case where the pattern probability distribution is almost uniform. For equally probable patterns, the linear term of the MPE variance vanishes, regardless of time scale. In this case, we need to increase the order of the approximation to the quadratic term. Hence, for MPE values close to the maximum, the variance presents a quadratic growth respect to scale.

We found the MPE variance estimator obtained in this article to be almost efficient. When the time scale is small compared to the signal length, the MPE variance resembles the MPE CRLB closely. Since the CRLB is equal to the first term on the Taylor series approximation for the variance, this implies that the efficiency limit also changes with the MPE measurement itself. Since this effect also comes purely from the pattern distribution, we cannot correct it with the established improvements of the MPE algorithm, like Composite MPE or Refined Composite MPE [13].

By knowing the variance of the MPE, we can improve the interpretation of this estimator in real-life applications. This is important because researchers can impose a precision criterion over the MPE measurements, given the characteristics of the data to analyze. For example, the electrical load behavior analysis can be achieved using short-term measurements where the time scale is a limiting factor for MPE. By knowing the variance, we can compute the maximum time scale until which the MPE is still significant.

By better understanding the advantages and disadvantages of the MPE technique, it is possible to have a clear benchmark for statistically significant comparisons between signals at any time scale.

**Author Contributions:** Conceptualization, D.A. and J.M.; methodology, D.A.; software, D.A.; validation, J.M., R.P. and B.O.; formal analysis, D.A.; resources, D.A.; writing—original draft preparation, D.A.; writing—review and editing, J.M. and R.P.; visualization, D.A.; supervision, B.O.; project administration, B.O..

**Funding:** The work is part of the ECCO project supported by the french Centre Val-de-Loire region under the contract #15088PRI.

**Acknowledgments:** This work was supported by CONACyT, Mexico, with scholarship 439625.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

PE	Permutation Entropy
MSE	Multiscale Entropy
MPE	Multiscale Permutation Entropy
CRLB	Cramér–Rao Lower Bound

## Appendix A. Multinomial Moment Matrices

In this section we will briefly derive the expressions for the Covariance, Coskewness, and Cokurtosis matrices for a multinomial distribution. These will be necessary in the calculation of the MPE variance in Section 2.2.

First, we start with a multinomial random variable with the following characteristics,

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} Y_1 \\ \vdots \\ Y_{d!} \end{bmatrix} = \begin{bmatrix} np_1 + \Delta Y_1 \\ \vdots \\ np_{d!} + \Delta Y_{d!} \end{bmatrix} = n\mathbf{p} + \Delta\mathbf{Y} \sim \text{Mu}(n, p_1, \dots, p_{d!}) \\ \hat{\mathbf{p}} &= \frac{1}{n}\mathbf{Y} = \mathbf{p} + \frac{1}{n}\Delta\mathbf{Y} \end{aligned} \quad (\text{A1})$$

where we use the same definitions as in Equations (10) and (12), where  $\mathbf{p}$  is the vector composed of all pattern probabilities  $p_i$ , and  $\hat{\mathbf{p}}$  is the estimation of  $\mathbf{p}$ . We will also use the definitions in Section 2.2, where  $d!$  is the number of possible patterns (number of events in the sample space),  $m$  is the MPE time scale, and  $N$  is the length of the signal. For the remainder of this section, we will assume  $m$  and  $N$  are such that  $n = \frac{N}{m}$  is integer.

We note that  $Y$  is composed of a deterministic part  $np_i$ , and a random part  $\Delta Y_i$ . It should be evident that  $E[\Delta Y_i] = 0$  and  $\Delta Y_i$  is identically distributed to  $Y$ .

We proceed to calculate the Covariance matrix of the binomial random variable  $\Delta\mathbf{Y}$ . We know that

$$E[\Delta Y_i^2] = np_i(1 - p_i) \quad (\text{A2})$$

$$E[\Delta Y_i \Delta Y_j] = -np_i p_j \quad (\text{A3})$$

for  $i = 1, \dots, d!$  and  $j = 1, \dots, d!$ . Thus, if we gather all possible combinations of  $i$  and  $j$  in the Covariance matrix, we get

$$E[\Delta\mathbf{Y}\Delta\mathbf{Y}^T] = n(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T) \quad (\text{A4})$$

Similarly, the skewness and coskewness can be expressed as,

$$E[\Delta Y_i^3] = 2np_i^3 - 3np_i^2 + np_i \quad (\text{A5})$$

$$E[\Delta Y_i^2 \Delta Y_j] = 2np_i^2 p_j - np_i p_j \quad (\text{A6})$$

which yields to the Coskewness matrix

$$E \left[ \Delta Y (\Delta Y^{\circ 2})^T \right] = 2n \left( \mathbf{p}^{\circ 2} \mathbf{p}^T - \text{diag}(\mathbf{p}^{\circ 2}) \right) + n(\text{diag}(\mathbf{p}) - \mathbf{p} \mathbf{p}^T) \tag{A7}$$

where we use again the vector definitions in Equation (12).

Lastly, we follow the same procedure to obtain the Cokurtosis matrix, by obtaining the values

$$E \left[ \Delta Y_i^4 \right] = 3n(n-2)p_i^4 - 6n(n-2)p_i^3 + n(3n-7)p_i^3 + np_i \tag{A8}$$

$$E \left[ \Delta Y_i^2 \Delta Y_j^2 \right] = 3n(n-2)p_i^2 p_j^2 - n(n-2)(p_i^2 p_j + p_i p_j^2) + n(n-1)p_i p_j \tag{A9}$$

which combines in the matrix as follows

$$E[\Delta Y^{\circ 2} (\Delta Y^{\circ 2})^T] = 3n(n-2) \mathbf{p}^{\circ 2} (\mathbf{p}^{\circ 2})^T - n(n-2) (\mathbf{p}^{\circ 2} \mathbf{p}^T + \mathbf{p} (\mathbf{p}^{\circ 2})^T) + n^2 \mathbf{p} \mathbf{p}^T - 4n(n-2) \text{diag}(\mathbf{p}^{\circ 3}) + 2n(n-3) \text{diag}(\mathbf{p}^{\circ 2}) + n(\text{diag}(\mathbf{p}) - \mathbf{p} \mathbf{p}^T). \tag{A10}$$

By taking advantage of this expressions, we are able to calculate the MPE variance in Section 2.2.

### Appendix B. Cramér–Rao Lower Bound of MPE

In this section we will explicitly develop the calculations of  $CRLB(H)$ . Before calculating the elements of the Information matrix, we should note that the parameter vector  $\mathbf{p}$  in Equation (12) has  $d! - 1$  degrees of freedom. since the elements of  $\mathbf{p}$  represent the probabilities of each possible pattern in the sample space, the sum of the elements of  $\mathbf{p}$  must be one. We are subject to the restriction

$$\sum_{i=1}^{d!} p_i = 1$$

$$p_{d!} = 1 - \sum_{i=1}^{d!-1} p_i \tag{A11}$$

We should note that, as a consequence of this restriction,  $p_{d!}$  is not an independent variable. We will define a new parameter vector  $\mathbf{p}_*$ , such that

$$\mathbf{p}_* = [p_1, \dots, p_{d!-1}]^T \tag{A12}$$

which has the same degrees of freedom as  $\mathbf{p}$ . This will be necessary to use the lemma (A19) as we shall see later in this section.

Now, we calculate all the partial derivatives of  $H$  (4) respect to the pattern probabilities  $\mathbf{p}$ , such that, for  $j = 1, \dots, (d! - 1)$

$$H = - \sum_{i=1}^{d!} p_i \ln p_i = - \sum_{i=1}^{d!-1} p_i \ln p_i - p_{d!} \ln p_{d!}$$

$$\frac{\partial H}{\partial p_j} = \ln p_{d!} - \ln p_j$$

$$\frac{\partial H}{\partial \mathbf{p}_*} = \ln p_{d!} \mathbf{1} - \mathbf{I}_* \tag{A13}$$

where

$$\mathbf{1} = [1, \dots, 1]^T \tag{A14}$$

$$\mathbf{I}_* = [\ln(p_1), \dots, \ln(p_{d!-1})]^T \tag{A15}$$

following the same logic as Equation (A12).

The next step is to compute the Fisher Information matrix. We know from Equation (6) that  $\hat{\mathbf{p}}$ , the pattern probabilities estimated from a signal, are multinomially distributed. The number of trials in the multinomial variable is  $N/m$ .

$$f_{\hat{\mathbf{p}}}(\hat{\mathbf{p}}; \mathbf{p}) = f_Y(\mathbf{y}; \mathbf{p}) = \frac{N!}{m!} \prod_{i=1}^{d!} \frac{p_i^{y_i}}{y_i!} = \frac{N!}{m!} \frac{p_{d!}^{y_{d!}}}{y_{d!}!} \prod_{i=1}^{d!-1} \frac{p_i^{y_i}}{y_i!} \tag{A16}$$

where

$$\mathbf{y} = [y_1, \dots, y_{d!}]^T \tag{A17}$$

is the count of patterns in the signal, as described in Equations (1) and (6). Since  $\hat{\mathbf{p}} = \frac{m}{N} Y$ , they have the same probability distribution. Using (A16),  $\mathbf{I}(\mathbf{p}_*)$  is calculated as follows,

$$\begin{aligned} \ln(f_Y(\mathbf{y}; \mathbf{p})) &= \ln\left(\frac{N!}{m!}\right) + y_{d!} \ln(p_{d!}) + \sum_{i=1}^{d!-1} y_i \ln(p_i) - \ln(y_{d!}!) - \sum_{i=1}^{d!-1} \ln(y_i!) \\ \frac{\partial \ln(f_Y)}{\partial p_j} &= y_j p_j^{-1} - y_{d!} p_{d!}^{-1} \\ \frac{\partial^2 \ln(f_Y)}{\partial p_j^2} &= -y_j p_j^{-2} - y_{d!} p_{d!}^{-2} \\ \frac{\partial^2 \ln(f_Y)}{\partial p_j \partial p_k} &= -y_{d!} p_{d!}^{-2} \\ -E \left[ \frac{\partial^2 \ln(f_Y)}{\partial p_j^2} \right] &= \frac{N}{m} p_j^{-1} + \frac{N}{m} p_{d!}^{-1} \\ -E \left[ \frac{\partial^2 \ln(f_Y)}{\partial p_j \partial p_k} \right] &= \frac{N}{m} p_{d!}^{-1} \\ \mathbf{I}(\mathbf{p}_*) &= \frac{N}{m} \left( \text{diag}(\mathbf{p}_*^{\circ-1}) + p_{d!}^{-1} \mathbf{1} \cdot \mathbf{1}^T \right) \end{aligned} \tag{A18}$$

where  $\mathbf{1} \cdot \mathbf{1}^T$  is a square matrix of ones, with rank 1. We should note that  $j = 1, \dots, (d! - 1)$  and  $k = 1, \dots, (d! - 1)$ , so  $\mathbf{I}(\mathbf{p}_*)$  is of size  $(d! - 1) \times (d! - 1)$ . Because of the constraint in (A11),  $p_{d!}$  offers no additional information. Moreover, we guarantee that  $\mathbf{I}(\mathbf{p}_*)$  is non-singular, and thus, invertible.

To get  $\text{CRLB}(\mathbf{p}_*)$ , we will need the inverse of  $\mathbf{I}(\mathbf{p}_*)$ : We will use the following lemma from [19]. If  $A$  and  $A + B$  are non-singular matrices, and  $B$  has rank 1, then,

$$(A + B)^{-1} = A^{-1} - \frac{1}{1 + \text{tr}(BA^{-1})} A^{-1} B A^{-1} \tag{A19}$$

therefore

$$\begin{aligned} \mathbf{I}(\mathbf{p}_*)^{-1} &= \frac{m}{N} (\text{diag}(\mathbf{p}_*^{\circ-1}) + p_{d!}^{-1} \mathbf{1} \cdot \mathbf{1}^T)^{-1} \\ &= \frac{m}{N} \left( \text{diag}(\mathbf{p}_*) - \frac{p_{d!}^{-1}}{1 + p_{d!}^{-1}(1 - p_{d!})} \mathbf{p}_* \mathbf{p}_*^T \right) \\ \mathbf{I}(\mathbf{p}_*)^{-1} &= \frac{m}{N} \left( \text{diag}(\mathbf{p}_*) - \mathbf{p}_* \mathbf{p}_*^T \right) = \text{CRLB}(\mathbf{p}_*). \end{aligned} \tag{A20}$$

Lastly, if we introduce Equations (A13) and (A20) into (23), we get

$$\begin{aligned} \text{CRLB}(H) &= \frac{m}{N} (\ln(p_{d!}) \mathbf{1} - \mathbf{l}_*)^T \left( \text{diag}(\mathbf{p}_*) - \mathbf{p}_* \mathbf{p}_*^T \right) (\ln(p_{d!}) \mathbf{1} - \mathbf{l}_*) \\ &= \frac{m}{N} (\ln^2(p_{d!}) \mathbf{1}^T \text{diag}(\mathbf{p}_*) \mathbf{1} - \ln^2(p_{d!}) \mathbf{1}^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{1} + \mathbf{l}_*^T \text{diag}(\mathbf{p}_*) \mathbf{l}_* - \mathbf{l}_*^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{l}_* \\ &\quad - 2 \ln(p_{d!}) \mathbf{1}^T \text{diag}(\mathbf{p}_*) \mathbf{l}_* + 2 \ln(p_{d!}) \mathbf{1}^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{l}_*). \end{aligned} \tag{A21}$$

By noting the following relations,

$$\begin{aligned}
 \mathbf{1}^T \text{diag}(\mathbf{p}_*) \mathbf{1} &= \sum_{i=1}^{d!-1} p_i = 1 - p_{d!} \\
 \mathbf{1}^T \mathbf{p}_* &= \sum_{i=1}^{d!-1} p_i = 1 - p_{d!} \\
 \mathbf{1}^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{1} &= \left( \mathbf{1}^T \mathbf{p}_* \right)^2 = (1 - p_{d!})^2 \\
 \mathbf{l}_*^T \text{diag}(\mathbf{p}_*) \mathbf{l}_* &= \sum_{i=1}^{d!-1} p_i \ln^2 p_i \\
 \mathbf{l}_*^T \mathbf{p}_* &= \sum_{i=1}^{D-1} p_i \ln p_i = -H - p_{d!} \ln p_{d!} \\
 \mathbf{l}_*^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{l}_* &= \left( \mathbf{l}_*^T \mathbf{p}_* \right)^2 = H^2 + 2H p_{d!} \ln p_{d!} + p_{d!}^2 \ln^2 p_{d!} \\
 \mathbf{1}^T \text{diag}(\mathbf{p}_*) \mathbf{l}_* &= \mathbf{p}_*^T \mathbf{l}_* = -H - p_{d!} \ln p_{d!} \\
 \mathbf{1}^T \mathbf{p}_* \mathbf{p}_*^T \mathbf{l}_* &= -(1 - p_{d!})(H + p_{d!} \ln p_{d!}) \\
 &= -H - p_{d!} \ln p_{d!} + p_{d!} H + p_{d!}^2 \ln p_{d!}
 \end{aligned}$$

we can simplify and rewrite Equation (A21) as

$$\text{CRLB}(H) = \frac{m}{N} \left( \sum_{i=1}^{d!} p_i \ln^2(p_i) - H^2 \right) = \frac{m}{N} \mathbf{l}^T \boldsymbol{\Sigma}_p \mathbf{l}$$

which is what we obtained in Equation (27).

## References

- Shannon, C.E. A Mathematical Theory of Communication. *SIGMOBILE Mob. Comput. Commun. Rev.* **2001**, *5*, 3–55. [\[CrossRef\]](#)
- Goldberger, A.L.; Peng, C.K.; Lipsitz, L.A. What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* **2002**, *23*, 23–26. [\[CrossRef\]](#)
- Cashaback, J.G.A.; Cluff, T.; Potvin, J.R. Muscle fatigue and contraction intensity modulates the complexity of surface electromyography. *J. Electromyogr. Kinesiol.* **2013**, *23*, 78–83. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wu, Y.; Chen, P.; Luo, X.; Wu, M.; Liao, L.; Yang, S.; Rangayyan, R.M. Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters. *Biomed. Signal Process. Control* **2017**, *31*, 265–271. [\[CrossRef\]](#)
- Aquino, A.L.L.; Ramos, H.S.; Frery, A.C.; Viana, L.P.; Cavalcante, T.S.G.; Rosso O.A. Characterization of electric load with Information Theory quantifiers. *PHYSICA A* **2017**, *465*, 277–284. [\[CrossRef\]](#)
- Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, C.; Li, K.; Zhao, L.; Liu, F.; Zheng, D.; Liu, C.; Liu, S. Analysis of heart rate variability using fuzzy measure entropy. *Comput. Biol. Med.* **2013**, *43*, 100–108. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [\[CrossRef\]](#) [\[PubMed\]](#)
- Aziz, W.; Arif, M. Multiscale Permutation Entropy of Physiological Time Series. In Proceedings of the 2005 Pakistan Section Multitopic Conference, Karachi, Pakistan, 24–25 December 2005; pp. 1–6. [\[CrossRef\]](#)
- Costa, M.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Physica A* **2003**, *330*, 53–60. [\[CrossRef\]](#)

12. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **2012**, *14*, 1553–1577. [[CrossRef](#)]
13. Humeau-Heurtier, A.; Wu, C.W.; Wu, S.D. Refined Composite Multiscale Permutation Entropy to Overcome Multiscale Permutation Entropy Length Dependence. *IEEE Signal Process. Lett.* **2015**, *22*, 2364–2367. [[CrossRef](#)]
14. Bandt, C.; Shiha, F. Order Patterns in Time Series. *J. Time Ser. Anal.* **2007**, *28*, 646–665. [[CrossRef](#)]
15. Zunino, L.; Pérez, D.G.; Martín, M.T.; Garavaglia, M.; Plastino, A.; Rosso, O.A. Permutation entropy of fractional Brownian motion and fractional Gaussian noise. *Phys. Lett. A* **2008**, *372*, 4768–4774. [[CrossRef](#)]
16. Little, D.J.; Kane, D.M. Permutation entropy of finite-length white-noise time series. *Phys. Rev. E* **2016**, *94*, 022118. [[CrossRef](#)] [[PubMed](#)]
17. Dávalos, A.; Jabloun, M.; Ravier, P.; Buttelli, O. Theoretical Study of Multiscale Permutation Entropy on Finite-Length Fractional Gaussian Noise. In Proceedings of the 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1092–1096.
18. Friedlander, B.; Francos, J.M. Estimation of Amplitude and Phase Parameters of Multicomponent Signals. *IEEE Trans. Signal Process.* **1995**, *43*, 917–926. [[CrossRef](#)]
19. Miller, K.S. On the Inverse of the Sum of Matrices. *Math. Mag.* **1981**, *54*, 67–72. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).