Entropy Analysis for the Climate Data and the Spring-mass Data

Release 0.9

G-H Gweon and H-S Lee

Apr 02, 2025

CONTENTS

CHAPTER

ONE

INTRODUCTION

This note summarizes some developing thoughts on analyzing data that involves a student exploring a multi-parameter space.

CHAPTER

TWO

ENTROPY, CORRELATION, OR SOMETHING ELSE

2.1 Why?

What is the sign that a student is learning fruitfully in a given learning activity in which a set of parameters can be varied by the student?

One might rule out an effective learning if all knobs available are turned in a chaotic manner.

One might also rule out an effective learning if only one knob is turned and it is turned only by a minute amount or too few times.

If the pattern involves some initial exploration of all knobs, followed by a settling down on one knob, which is explored extensively and systematically, then this might be recognized easily as a great learning pattern by a human.

The question that we ask here is—can we develop a metric, or a set of metrics, that will automatically summarize these differences between these and other different learning patterns?

2.2 Entropy

Suppose that a student explores one axis in the full parameter space (which is four dimensional in the climate game).

1. The parameter variation is systematic.
2. The total range of the parameter variation is reasonably large.
3. The number of trials is reasonably large.

> **Note**
>
> How large is "reasonably large"? A priori, we may or may not know the answer. In the latter case, maybe we can determine the answer from the data, choosing something like "greater than the median."

Entropy
: A measurement of the disorderliness of the whole set of data points. Too large an entropy is obviously bad. But, too small an entropy is also bad, probably, in some sense, which may or may not be captured by the analysis.

Correlation
: A measurement of the point to point relation.

CHAPTER

THREE

VARIOUS ENTROPY CONCEPTS IN THE LITERATURE

3.1 Gibbs entropy

$$S_G = -k_B \sum_i p_i \log p_i$$

This is the thermodynamic entropy of an ensemble distribution, where each state in the ensemble is denoted by the state index i.

> **Note**
>
> log without any subscript stands for the natural logarithm in this note.

Of course, this becomes the so-called Boltzmann entropy, $k_B \log \Gamma$ where Γ is the number of available microstates in the micro-canonical ensemble case.

3.2 Shannon entropy

Here, we can consider a certain set of words, each of which can be represented in a fixed number M of bits. A random word can be assigned an index i, where $0 \leq i < N = 2^M$. The word distribution can be described as a probability distribution p_i, given by n_i/N, where n_i is the number of occurrences of word i.

$$S_S = -\sum_i p_i \log_2 p_i$$

defines the Shannon entropy, and gives the numbers of bits to express all words in the set. So, one may call this a measure of the information content.

3.3 Tsallis entropy

This is a formulation of a non-additive entropy.

$$S_T = \frac{1 - \sum_i p_i^q}{q - 1}$$

where the parameter q defines the non-additivity of the entropy. Namely, if $p(A, B) = p(A)p(B)$ (two sets A and B are independent), then the Tsallis entropy is not additive. Rather it is given by, as can be shown easily,

$$S_T(A \cup B) = S_T(A) + S_T(B) + (q - 1)S_T(A)S_T(B).$$

If $q \to 1$, then $S_T \to S_S/\log 2 = S_G/k_B$.

Tsallis entropy is an interesting concept, and is expected to be very relevant to complex systems. For now, though, we shall limit ourselves to the Shannon entropy.

3.4 Entropy—some general properties

From here on, we will use $\log$ instead of $\log_2$ to denote the Shannon entropy. Also, we shall use S without any subscript to mean the Shannon (or Gibbs) entropy.

Consider the joint entropy

$$S_N = S(X_1, ..., X_N) = -\sum_{X_1,...,X_N} p(x_1, ..., x_N) \log p(x_1, ..., x_N)$$

where x_i is a random variable and X_i is the corresponding set. Consider for example two variables, x and y.

$$\begin{aligned} S(X,Y) &= -\textstyle\sum_{x,y} p(x,y) \log p(x,y), \\ &= -\textstyle\sum_{x,y} p(x|y)p(y) \log \left[p(x|y)p(y)\right], \\ &= -\textstyle\sum_{y}\sum_{x} p(y)p(x|y) \log p(x|y) - \sum_{y} p(y) \log p(y). \end{aligned}$$

The first term is the so-called conditional entropy, $S(X|Y)$, while the second term is just the entropy of a single variable $S(Y)$. Now, by substituting $X = X_N$ and $Y = X_1, ..., X_{N-1}$, it is clear that we get

$$S_N = S(X_1, ..., X_N) = \sum_{i=1}^{N} S(X_i|X_{i-1}, X_{i-2}, ..., X_1)$$

It stands to reason that as more random variable is added the entropy increases.

From now on, let us interpret the indices $1, ..., N$ as time indices. That is, we consider a stochastic process.

The **entropy rate** of a stochastic process is defined as

$$s(\vec{X}) = \lim_{N\to\infty} \frac{S_N}{N}.$$

Also, define **the conditional entropy limit**

$$S_c(X) = \lim_{N\to\infty} S(X_N|X_{N-1}, ..., X_1).$$

A stochastic process is referred to as a **stationary process**, if the joint probability distribution is invariant upon the translation in the time index, i.e,

$$p(X_1 = x_1, ..., X_N = x_n) = p(X_{1+l} = x_1, ..., X_{N+l} = x_n).$$

It is now easy to see that, for a stationary process, S_c is a well-defined convergent quantity, since

$$S(X_N|X_{N-1}, ..., X_1) \leq S(X_N|X_{N-1}, ..., X_2) = S(X_{N-1}|X_{N-2}, ..., X_1),$$

where the inequality is due to less conditioning (the difference is the negative of the mutual entropy, $I(X,Y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$, which is known to be non-negative due to the "Jensen's inequality") and the equality is due to the stationary process.

If $S_c(X)$ is well-defined (as for a stationary process), then it follows that

$$S_c(X) = s(\vec{X}),$$

since

$$S_N \to NS_c \quad \text{as} \quad N \to \infty.$$

Also, it follows that

$$S_{N+1} - S_N \to S_c(X) = s(\vec{X}).$$

3.5 Kolmogorov entropy

The so-called Kolmogorov or Kolmogorov-Sinai entropy originated from the chaos theory and, thus, it concerns the entropy of a dynamical process. The essence of this entropy is summarized well in here [mathworld-Kolmogorov-Entropy]. As much (or little) as the chaos theory is capable of explaining the thermodynamics entropy, this entropy is expected to become Gibbs entropy (or not).

Let us use the language of the classical mechanics in physics, which can be, of course, generalized to the language of an abstract mathematical mapping, if one wishes.

In the Hamiltonian formulation of the Newtonian mechanics, the time evolution of a Newtonian system is represented by a mapping in the phase space. Let D be the dimension of the phase space. For example, if there are five particles in three spatial dimensions, then $D = 5 \times 2 \times 3 = 30$, since the phase space is built by attaching a momentum axis to any spatial axis (thus 2×3 is the dimension of the phase space for a single particle in three physical dimensions).

Now consider dividing the phase space into vary small elements, each of which has a linear dimension ϵ along any axis: so each volume element is given by ϵ^D. Each volume element can be given an index i. Define

$$S_n = -\sum_{i_0,\ldots,i_n} p_{i_0,\ldots,i_n} \log p_{i_0,\ldots,i_n}$$

where n is the time step index (time $t = n\tau$, where τ is a small time increment—see below), and $p_{i_0,\ldots,i_n}$ is the probability that each of the string of volume elements $i_0, \ldots, i_n$ contains a trajectory.

The Kolmogorov entropy is then defined as

$$S_K = \lim_{\tau\to 0}\lim_{\epsilon\to 0}\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}(S_{n+1} - S_n)\,.$$

Note that $S_{n+1} - S_n$ is the number of additional bits required, going from step n to the next step, to describe the motion of all states in the ensemble. Since the sum is over adjacent differences, we get

$$S_K = \lim_{\tau\to 0}\lim_{\epsilon\to 0}\lim_{N\to\infty} \frac{S_N - S_0}{N} = \lim_{\tau\to 0}\lim_{\epsilon\to 0}\lim_{N\to\infty} \frac{S_N}{N}.$$

So, in the language of the previous section, the Kolmogorov entropy is the entropy rate, which is equal to the conditional entropy limit, assuming that the latter exists.

3.6 Approximate entropy (practical entropy)

While the Kolmogorov entropy is well-defined mathematically, and it would be computable for a complex system with a bit of physics constraints, it is not readily computable for practical use. Therefore, some practical entropy concepts have been invented.

The concept of approximate entropy has been proposed by [Pincus] as a practical way to compute a quantity that approximates the Kolmogorov entropy.

Here, we consider a one-dimensional time series, consisting of N numbers

$$u_0, u_1, u_1, \ldots, u_n, \ldots, u_{N-1}.$$

We choose a positive integer m and a positive real number r. If the difference between two adjacent value of u is less than r, then we will designate them to be *the same*. So, r plays the role of ϵ in the definition of the Kolmogorov entropy, except that here we do not take the zero limit.

> **Note**
>
> In this section and the following section, "the same" means "the same within tolerance r."

Form $N-m+1$ vectors, $\vec{x}_{i,m}$, as follows.

$$\begin{aligned}\vec{x}_{0,m} &= (u_0,\ldots,u_{m-1}),\\ &\ldots\\ \vec{x}_{i,m} &= (u_i,\ldots,u_{m+i-1}),\\ &\ldots\\ \vec{x}_{N-m,m} &= (u_{N-m},\ldots,u_{N-1}).\end{aligned}$$

Now, we have a mapping in an m-dimensional space. For a given $\vec{x}_{i,m}$, we can count the number of vectors, $\vec{x}_{j,m}$, that are *the same* as $\vec{x}_{i,m}$, i.e., if the distance between them is less than r. The distance between vectors can be defined by a suitable metric such as the maximum of the component-wise differences. Let the number of vectors, $\vec{x}_{j,m}$, that are the same as $\vec{x}_{i,m}$ be $E_{i,m}$. Then, define

$$\begin{aligned}C_{i,m} &= \tfrac{E_{i,m}}{N-m+1},\\ \Phi_m(r) &= \tfrac{1}{N-m+1}\sum_{i=0}^{N-m}\log C_{i,m}.\end{aligned}$$

The probability $C_{i,m}$ is the probability that a randomly chosen vector $\vec{x}_{j,m}$ is the same as $\vec{x}_{i,m}$. Note that $j=i$ is included in this consideration. In particular, $C_{i,m}$ is guaranteed to be positive since $E_{i,m}\geq 1$ for the inclusion of the self-match.

The approximate entropy is defined as

$$S_{ap}=\Phi_m(r)-\Phi_{m+1}(r).$$

Since $S_{ap}=\frac{1}{N-m+1}\sum_{i=0}^{N-m}\log C_{i,m}-\frac{1}{N-m}\sum_{i=0}^{N-m-1}\log C_{i,m+1}$, it is easy to see that, for large N, $S_{ap}\approx \frac{1}{N-m}\sum_{i=0}^{N-m-1}\Big(-\log\frac{E_{i,m+1}}{E_{i,m}}\Big)$. Here, $C_{i,m+1}/C_{i,m}(\approx E_{i,m+1}/E_{i,m})$ is the conditional probability that if a vector is the same as $\vec{x}_{i,m}$ it remains the same as $\vec{x}_{i,m+1}$ at the next time step on average. The approximate entropy is the average of the negative $\log$ of this conditional probability; so it is a highly convolved quantity, although it is clear that it codifies the complexity of the dynamics. Using this approximate formula, it is clear that, for large N, we get $S_{ap}\leq\log(N-m+1)\approx\log(N-m)$, since the entropy is maximized if $E_{i,m}=N-m+1$ and $E_{i,m+1}=1$.

More precisely speaking, we see that, for any value of N, $\Phi_m\leq 0$ and $\Phi_{m+1}\geq\log\frac{1}{N-m}$. And so, actually,

$$S_{ap}\leq\log(N-m)$$

for any value of N.

Is the approximate entropy non-negative? This is *not* the case. Consider the case then $E_{i,m}=1$. In this case, necessarily $E_{i,m+1}=1$ also. Then, we get $\Phi_m=-\log(N-m+1)$ and $\Phi_{m+1}=-\log(N-m)$. Therefore, $S_{ap}=\log\frac{N-m}{N-m+1}<0$. As $N\to\infty$, this value is given by $\approx-\frac{1}{N-m+1}$. Let us do some more analysis.

$$\begin{aligned}S_{ap} &= \frac{1}{N-m+1}\sum_{i=0}^{N-m}\log\frac{E_{i,m}}{N-m+1}-\frac{1}{N-m}\sum_{i=0}^{N-m-1}\log\frac{E_{i,m+1}}{N-m}\\ &= \frac{\log\frac{E_{N-m,m}}{N-m+1}}{N-m+1}+\frac{1}{N-m+1}\sum_{i=0}^{N-m-1}\log\frac{E_{i,m}}{N-m+1}-\frac{1}{N-m}\sum_{i=0}^{N-m-1}\log\frac{E_{i,m+1}}{N-m}\\ &\geq \frac{1}{N-m+1}\left(\log\frac{E_{N-m,m}}{N-m+1}+\sum_{i=0}^{N-m-1}\log\frac{E_{i,m}(N-m)}{E_{i,m+1}(N-m+1)}\right)\\ &\geq \frac{1}{N-m+1}\left(\log\frac{1}{N-m+1}+(N-m)\log\frac{N-m}{N-m+1}+\sum_{i=0}^{N-m-1}\log\frac{E_{i,m}}{E_{i,m+1}}\right)\\ &\geq \frac{1}{N-m+1}\left(\log\frac{1}{N-m+1}+(N-m)\log\frac{N-m}{N-m+1}\right)\\ &\geq \frac{1}{N-m+1}\Big((N-m)\log(N-m)-(N-m+1)\log(N-m+1)\Big).\end{aligned}$$

So, as $N \to \infty$, $S_{ap} \geq -\frac{\log(N-m)}{N-m+1}$. That is, the actual minimum value of S_{ap} may arise when $E_{i,m} \sim N-m$, not when $E_{i,m} = 1$, and $E_{i,m+1} = E_{i,m}$ or $E_{i,m} - 1$. However, going back to the first expression, it is easy to see that even in these cases, $S_{ap} \approx -\frac{O(1)}{N-m+1}$. So, the factor $\log(N-m)$ in the numerator arises from the first inequality estimation step in the above analysis, where $1/(N-m)$ is replaced by $1/(N-m+1)$. So, the minimum of S_{ap} seems to be $-\frac{O(1)}{N-m+1}$, rather than $-\frac{\log(N-m)}{N-m+1}$. So,

$$-\frac{O(1)}{N-m+1} \leq S_{ap} \leq \log(N-m).$$

3.7 Sample entropy (practical entropy)

Another practical entropy concept is the sample entropy.

This entropy has been suggested by [Richman-and-Moorman] as an improvement over S_{ap} . In this work, Richman and Moorman argue successfully that the self-match ($E_{i,m} \geq 1$) included in the definition of S_{ap} gives it a bias. Namely, S_{ap} is always less than the correct value, which can be obtained only in the limit of zero r and infinite values of N and m. They show that the systematic error of the practical entropy due to this bias becomes pronounced when a small r value is used.

The strategy that they take to correct this bias is the following.

1. Do not consider the $j = i$ case when counting vectors that are the same as $\vec{x}_{i,m}$. So, now, $E_{i,m}$ can be zero.
2. Define the sample entropy as

$$S_{sa} = -\log\left(\frac{A_m}{B_m}\right) = -\log\left(\frac{\frac{1}{N-m}\sum_{i=0}^{N-m-1} A_{i,m}}{\frac{1}{N-m}\sum_{i=0}^{N-m-1} B_{i,m}}\right) = -\log\left(\frac{\sum_{i=0}^{N-m-1} E_{i,m+1}}{\sum_{i=0}^{N-m-1} E_{i,m}}\right).$$

Here, there are subtle changes in the summation indices compared to those of the approximate entropy. *Both* in the numerator and in the denominator, the index i goes from 0 to $N-m-1$, counting only the first $N-m$ vectors whether we are considering the m case (denominator) or the $m+1$ case (numerator). The symbols $A_{i,m}$ and $B_{i,m}$ are defined as

$$A_{i,m} = \frac{E_{i,m+1}}{N-m-1},$$
$$B_{i,m} = \frac{E_{i,m}}{N-m-1}.$$

So, $B_{i,m}$ is the probability that any $\vec{x}_{j\neq i,m}$ is the same as $\vec{x}_{i,m}$. Similarly, $A_{i,m}$ is the probability that any $\vec{x}_{j\neq i,m+1}$ is the same as $\vec{x}_{i,m+1}$ for $j = 0$ through $N-m-1$.

B_m (A_m) is the probability that two sequences will match m ($m+1$) points. And so, the sample entropy is sort of like the approximate entropy except that (i) logarithm is taken outside all sums, and (ii) the counting parameter $E_{i,m}$ avoids self-matches.

It turns out that the sample entropy does well for small r values, while at the same time it suffers a problem of no value soon if r becomes too small (see Fig. 2 of their paper). For instance, with $N = 100$ and $m = 2$, the two are basically the same if $0.5 \leq r \leq 1$ but S_{sa} does much better if $0.2 \leq r \leq 0.5$. If $r \leq 0.2$, S_{sa} cannot be defined (because A_m or B_m can be zero), while S_{ap} gives a strongly biased incorrect result.

Figure 1 of [Costa-PRB] has a nice simple explanation for how to calculate the sample entropy for a real data set.

Does the sample entropy have any upper maximum? Strictly speaking, the answer is no, since $S_{sa} \to +\infty$ if $E_{i,m+1} = 0$ while $\sum_i E_{i,m} > 0$. However, if we limit ourselves to cases when the sample entropy is well-defined, that is, if we do not consider cases where the numerator $\sum_i E_{i,m+1}$ or the denominator $\sum_i E_{i,m}$ is zero, then we see that the sample entropy has maximum possible value of $2\log(N-m)$ (when $\sum_i E_{i,m+1} = 1$ and $\sum_i E_{i,m} = (N-m)(N-m)$). So,

$$0 \leq S_{sa} \leq 2\log(N-m).$$

3.8 Renyi entropy

According to [Costa-PRB], the difference between S_{ap} and S_{sa} can be related to the Renyi entropy,

$$S_R(q) = \frac{\log(\sum_i p_i^q)}{1-q}.$$

According to [Costa-PRB] (and its reference 21, which is not available by UC lib), S_{ap} approximates the Renyi entropy of order $q = 1$, while S_{sa} approximates the Renyi entropy of order $q = 2$.

All this sounds a bit cryptic to me at this point, since I do not have reference 21 in my hand and since [Costa-PRB] discusses this point only in passing.

In any case, note that this Renyi entropy looks somewhat reminiscent of the Tsallis entropy above. They have in common that when $q \to 1$ they converge to the Shannon entropy.

3.9 Multi-scale entropy analysis

In this work [Costa-PRL], a simple coarse-graining idea is used to calculate the entropy as a function of coarse-graining scale. That is, the raw data u_i are converted to coarse-grained data y_i with coarse-graining scale ξ.

$$y_{\xi,i} = \frac{1}{\xi} \sum_{j=i\xi}^{(i+1)\xi-1} u_j.$$

In this equation, $j = 0, ..., N-1$, while $i = 0, ..., (N/\xi) - 1$.

Applied to the heartbeat data, the multi-scale entropy analysis is shown to be able to discern an unhealthy heart condition from a healthy heart condition only if the large scale correlation is included in the analysis by use of a large coarse-graining scale. It is demonstrated that the analysis with no coarse-graining can lead to a completely incorrect result (Figure 3 of [Costa-PRL]).

CHAPTER

FOUR

REFERENCES AND LINKS

CHAPTER

FIVE

INDICES AND OTHERS

- genindex
- pdf file
- search

BIBLIOGRAPHY

[mathworld-Kolmogorov-Entropy] *Kolmogorov entropy*, mathworld.Wolfram.com, http://mathworld.wolfram.com/KolmogorovEntropy.html

[Pincus] *Approximate entropy as a measure of system complexity*, S M Pincus, http://www.pnas.org/content/88/6/2297.

[Richman-and-Moorman] *Physiological time-series analysis using approximate entropy and sample entropy*, Joshua S. Richman, J. Randall Moorman, American Journal of Physiology - Heart and Circulatory Physiology Published 1 June 2000 Vol. 278 no. 6, H2039-H2049, http://ajpheart.physiology.org/content/278/6/H2039

[Costa-PRL] *Multiscale Entropy Analysis of Complex Physiologic Time Series*, Madalena Costa, Ary L. Goldberger, and C.-K. Peng Phys. Rev. Lett. 89, 068102 – Published 19 July 2002, http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.89.068102

[Costa-PRB] *Multiscale entropy analysis of biological signals*, Madalena Costa, Ary L. Goldberger, and C.-K. Peng Phys. Rev. E 71, 021906 – Published 18 February 2005, http://journals.aps.org/pre/abstract/10.1103/PhysRevE.71.021906

INDEX