

Chapter 3

Entropy for the Complexity of Physiological Signal Dynamics

Xiaohua Douglas Zhang

Abstract Recently, the rapid development of large data storage technologies, mobile network technology, and portable medical devices makes it possible to measure, record, store, and track analysis of biological dynamics. Portable noninvasive medical devices are crucial to capture individual characteristics of biological dynamics. The wearable noninvasive medical devices and the analysis/management of related digital medical data will revolutionize the management and treatment of diseases, subsequently resulting in the establishment of a new healthcare system. One of the key features that can be extracted from the data obtained by wearable noninvasive medical device is the complexity of physiological signals, which can be represented by entropy of biological dynamics contained in the physiological signals measured by these continuous monitoring medical devices. Thus, in this chapter I present the major concepts of entropy that are commonly used to measure the complexity of biological dynamics. The concepts include Shannon entropy, Kolmogorov entropy, Renyi entropy, approximate entropy, sample entropy, and multiscale entropy. I also demonstrate an example of using entropy for the complexity of glucose dynamics.

Keywords High-throughput phenotyping • Entropy • Complexity • Wearable medical device • Continuous monitoring

3.1 Introduction

In this century, the rapid development of genomics biotechnologies, large data storage technologies, mobile network technology, and portable medical devices makes it possible to measure, record, store, and track analysis of the genome, physiological dynamics, and living environment and subsequently to reveal the differences and uniqueness of an individual. These technologies may allow us to reclassify the diseases for making diagnostic and therapeutic strategies more precisely tailored to individual patients, leading the birth of precision medicine and personalized medicine. Portable noninvasive medical devices are crucial to

X.D. Zhang (✉)

Faculty of Health Sciences, University of Macau, Taipa, Macau, China

e-mail: douglaszhang@umac.mo

capture individual characteristics of biological dynamics. In fact, the rapid development of wearable, mobile, automatic, continuous, high-throughput medical device for measuring human biological parameters heralds a new era [5] – high throughput phenotyping era. In this new era, the wearable noninvasive medical devices and the analysis/management of related digital medical data will revolutionize the management and treatment of diseases, subsequently resulting in the establishment of a new healthcare system including those for the treatment and care of respiratory patients.

One of the key features that can be extracted from the data obtained by the high-throughput medical device is the complexity of physiological signals. Thus, complexity studies in various major diseases including respiratory diseases are arising as an important method to analyze these continuous monitoring data measured by the noninvasive medical devices. The complexity of physiological signals can be represented by entropy of biological dynamics contained in the physiological signals measured by continuous monitoring medical devices.

The initial entropy is derived from and applied to the physics of thermodynamics and statistical physics. Clausius introduced the concept of entropy in the 1850s [1] and was the first one to enunciate the second law of thermodynamics by saying that “entropy always increases.” Boltzmann was the first to state the logarithmic connection between entropy and probability in 1886. In 1948, Shannon [12] proposed an entropy (later known as Shannon entropy) and a large number of applications in information science. The Kolmogorov entropy [8] and Renyi entropy [10], which are developed on the basis of Shannon’s entropy, are widely used in the nonlinear dynamics of the physical system. The entropy can be applied to the experimental data of biological dynamics, such as approximate entropy [9], sample entropy [11], and multiscale entropy [2, 14], to quantify the physiological signals in the physiological dynamic system, such as heart rate, airflow, pressure in airway, signal sound, and so on. Hence, in this chapter, I should describe the above major concepts of entropy, show their connections, and demonstrate an example of using entropy. The original concepts of entropy proposed by Clausius and Boltzmann are rarely used for biological dynamics directly. Thus, I start with Shannon entropy.

3.2 Shannon Entropy

For a discrete random variable Y with a probability mass function $p(Y)$, the entropy is defined as the expectation of a function of Y , $I(Y)$, where $I(Y) = -\log(p(Y))$. That is,

$$H(Y) = E\{I(Y)\} = E\{-\log(p(Y))\} \quad (3.1)$$

In information science, $I(Y)$ is the information content of Y , which is also a random variable. Suppose that the random variable Y has possible values $\{y_1, y_2, \dots, y_n\}$

and a corresponding probability mass function of $p_i = \Pr(Y = y_i)$. The entropy can be expanded as [12]

$$\begin{aligned} H(Y) &= E\{I(Y)\} = E\{-\log(p(Y))\} = \sum_{i=1}^n (p_i \times (-\log(p_i))) \\ &= - \sum_{i=1}^n (p_i \log p_i) \end{aligned} \quad (3.2)$$

This definition of entropy can be extended to a continuous random variable with a probability density function $f(Y)$ as follows:

$$H(Y) = E\{I(Y)\} = E\{-\log(f(Y))\} = - \int f(Y) \times \log(f(Y)) dY \quad (3.3)$$

The entropy defined on the discrete variable is most commonly used. Hence, we focus on the entropy defined on a discrete variable. The concept of entropy can be easily extended to multidimensional random variable.

Suppose there are two events, X and Y , in question with I possibilities for the first and J for the second. Let $p(x_i, y_j)$ be the probability of the joint occurrence of x_i for the first and y_j for the second. The marginal mass density functions of X and Y are $p_X(x_i) = \sum_{j=1}^J p(x_i, y_j)$ and $p_Y(y_j) = \sum_{i=1}^I p(x_i, y_j)$, respectively. The entropy of the joint event is [12]

$$\begin{aligned} H(X, Y) &= E_{X,Y}\{I(X, Y)\} = E_{X,Y}\{-\log(p(X, Y))\} \\ &= - \sum_{i=1}^I \sum_{j=1}^J (p(x_i, y_j) \log(p(x_i, y_j))) \end{aligned} \quad (3.4)$$

The entropy of X and Y are, respectively,

$$\begin{aligned} H(X) &= E_X\{-\log(p(X))\} = - \sum_{i=1}^I (p_X(x_i) \log(p_X(x_i))) \\ &= - \sum_{i=1}^I \sum_{j=1}^J (p(x_i, y_j) \log(p_X(x_i))) = - \sum_{i=1}^I \sum_{j=1}^J \left(p(x_i, y_j) \log \left(\sum_{j=1}^J p(x_i, y_j) \right) \right) \end{aligned}$$

$$\begin{aligned} H(Y) &= E_Y\{-\log(p(Y))\} = - \sum_{j=1}^J (p_Y(y_j) \log(p_Y(y_j))) \\ &= - \sum_{j=1}^J \sum_{i=1}^I (p(x_i, y_j) \log(p_Y(y_j))) = - \sum_{i=1}^I \sum_{j=1}^J (p(x_i, y_j) \log(p_Y(y_j))) \\ &= - \sum_{i=1}^I \sum_{j=1}^J \left(p(x_i, y_j) \log \left(\sum_{i=1}^I p(x_i, y_j) \right) \right) \end{aligned}$$

The conditional probability of $Y = y_j$ given $X = x_i$ is $p(Y|X) = \frac{p(x_i y_j)}{\sum_{j=1}^n p(x_i y_j)} = \frac{p(x_i y_j)}{p_X(x_i)}$

The conditional entropy of two discrete random variables X and Y is defined as [12]

$$\begin{aligned} H_X(Y) &= E_{X,Y}\{I(Y|X)\} = E_{X,Y}\{-\log(p(Y|X))\} \\ &= -\sum_{i=1}^I \sum_{j=1}^J \left(p(x_i, y_j) \log \frac{p(x_i, y_j)}{\sum_{j=1}^n p(x_i, y_j)} \right) \end{aligned} \quad (3.5)$$

Based on Shannon [12], the entropy has the following major properties for serving as a measure of choice or information:

- (i) $H = 0$ if and only if all the p_i but one are zero, this one having the value unity. Thus only when we are certain of the outcome does H vanish. Otherwise H is positive.
- (ii) For a given n , H is a maximum and equal to $\log n$ when all the p_i are equal (i.e., $\frac{1}{n}$). This is also intuitively the most uncertain situation.
- (iii) Any change toward equalization of the probabilities p_1, p_2, \dots, p_n increases H . Thus if $p_1 < p_2$ and we increase p_1 , decreasing p_2 an equal amount so that p_1 and p_2 are more nearly equal, then H increases. More generally, if we perform any “averaging” operation on the p_i of the form $p'_i = \sum_j a_{ij} p_j$ where $\sum_j a_{ij} = 1$ and all $a_{ij} \geq 0$, then H increases (except in the special case where this transformation amounts to no more than a permutation of the p_i with H of course remaining the same).
- (iv) It is easily shown that $H(X, Y) \leq H(X) + H(Y)$ because $p(x_i, y_j) \geq p_X(x_i)p_Y(y_j)$. The equality holds only if the events are independent (i.e., $p(x_i, y_j) = p_X(x_i)p_Y(y_j)$). The uncertainty of a joint event is less than or equal to the sum of the individual uncertainties.
- (v) It is easily shown that $H(X, Y) = H(X) + H_X(Y)$. The uncertainty (or entropy) of the joint event X and Y is the uncertainty of X plus the uncertainty of Y given X is known.
- (vi) $H(Y) \geq H_X(Y)$ because $H(X) + H(Y) \geq H(X, Y) = H(X) + H_X(Y)$. The uncertainty of Y is never increased by knowledge of X . It will be decreased unless X and Y are independent events, in which case it is not changed.

3.3 Conditional Entropy

From Shannon’s definition [12] leading to the relationship of $H(X, Y) = H(X) + H(Y|X)$, i.e.,

$$\begin{aligned}
H(X, Y) &= -\sum_{i=1}^m \sum_{j=1}^n \left(p(x_i, y_j) \log(p(x_i, y_j)) \right) \\
&= -\sum_{i=1}^m \sum_{j=1}^n \left(p(x_i, y_j) \log \left(\sum_{j=1}^n p(x_i, y_j) \times \frac{p(x_i, y_j)}{\sum_{j=1}^n p(x_i, y_j)} \right) \right) \\
&= -\sum_{i=1}^m \sum_{j=1}^n \left(p(x_i, y_j) \log \left(\sum_{j=1}^n p(x_i, y_j) \right) \right) \\
&\quad - \sum_{i=1}^m \sum_{j=1}^n \left(p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{\sum_{j=1}^n p(x_i, y_j)} \right) \right) = H(X) + H_X(Y)
\end{aligned}$$

we know that X is not fixed at x_i when the conditional entropy $H(Y|X)$ is not for one fixed x_i .

It is interesting to see that, if $Z_i = Y|(X = x_i)$ is treated as a random variable, the entropy defined on Z_i will be

$$\begin{aligned}
H(Z_i) &= H(Y|X = x_i) = E_{Z_i} \{I(Z_i)\} = E_{Z_i} \{-\log(p(Z_i))\} \\
&= -\sum_{j=1}^n \left(p(y_j|x_i) \log(p(y_j|x_i)) \right) \\
&= -\sum_{j=1}^n \left(\frac{p(x_i, y_j)}{\sum_{j=1}^n p(x_i, y_j)} \log \left(\frac{p(x_i, y_j)}{\sum_{j=1}^n p(x_i, y_j)} \right) \right)
\end{aligned}$$

which differs from the conditional entropy of two discrete random variable. Note, here X is fixed at x_i . In another word, the conditional entropy differs from the entropy of a conditional probability or event. The conditional entropy of Y given X measures the uncertainty of Y given we know X regardless of what the value of X is, whereas the entropy of Y given X measures the uncertainty of Y given X equals one of its specific values. Their relationship can be shown below:

$$\begin{aligned}
H_X(Y) &= -\sum_{i=1}^m \sum_{j=1}^n \left(p(x_i, y_j) \log(p(y_j|x_i)) \right) \\
&= -\sum_{i=1}^m \sum_{j=1}^n \left(p(x_i) p(y_j|x_i) \log p(y_j|x_i) \right) \\
&= -\sum_{i=1}^m \left(p(x_i) \sum_{j=1}^n p(y_j|x_i) \log p(y_j|x_i) \right) = -\sum_{i=1}^m \left(p(x_i) H(Z_i) \right)
\end{aligned}$$

That is,

$$H_X(Y) = - \sum_{i=1}^m (p(x_i)H(Z_i)) \quad (3.6)$$

Hence, the conditional entropy is a weighted or average entropy of conditional probabilities or events.

3.4 Renyi Entropy

In mathematics, Shannon entropy can be seen as a special case of Renyi entropy [10] which is introduced here. In general, for a discrete random variable Y with a probability mass function $p_i = \Pr(Y = y_i)$, the Renyi entropy of order q , where $q \geq 0, q \neq 1$, is

$$R_q(Y) = - \frac{1}{q-1} \log \sum_{i=1}^n p_i^q \quad (3.7)$$

The limit of Renyi entropy for $q \rightarrow 1$ is the Shannon entropy, that is,

$$\lim_{q \rightarrow 1} R_q(Y) = - \lim_{q \rightarrow 1} \frac{1}{q-1} \log \sum_{i=1}^n p_i^q = - \sum_{i=1}^n (p_i \log p_i) \quad (3.8)$$

Consequently, it can be defined $R_{q=1}(Y) = \lim_{q \rightarrow 1} R_q(Y)$. In such a case, the Shannon entropy is a special case of Renyi entropy with $q = 1$. It can be demonstrated that

$$R_{q_1}(Y) > R_{q_2}(Y) \text{ when } q_1 > q_2 \quad (3.9)$$

3.5 Kolmogorov Entropy

Kolmogorov entropy (also known as metric entropy) is originated from the field of dynamical systems, which is defined as follows [8]. Consider a dynamical system with a phase space. Divide the phase space into a set $N(\epsilon)$ of disjoint D -dimensional hypercubes of content ϵ^D . Let $p(i_0, i_1, i_2, \dots, i_N)$ be the probability that a trajectory is in hypercube i_0 at $t = 0$, hypercube i_1 at $t = \tau$, hypercube i_2 at $t = 2\tau \dots$, hypercube i_N at $t = N\tau$.

Kolmogorov entropy is then defined as

$$K = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} \left(\sum_{i_0, i_1, i_2, \dots, i_N} (p(i_0, i_1, i_2, \dots, i_N) \log p(i_0, i_1, i_2, \dots, i_N)) \right) \quad (3.10)$$

The above definition relies on quite some jargons such as phase space, hypercube, content, trajectory, and so on in physics that a statistician is usually unfamiliar.

A dynamical system is a description of a physical system that evolves over time. The system has many states, and all states are represented in the state space of the system. The state space is also termed as phase space. A path in the state space describes the dynamics of the dynamical system. The path is termed as trajectory.

Kolmogorov entropy is the theoretical basis for approximate entropy, sample entropy, and multiscale entropy that are commonly used in time series, which relies on the concept that Kolmogorov entropy is the rate of change of Shannon entropy of a system. The Kolmogorov-Sinai entropy measures unpredictability of a dynamical system. The higher the unpredictability, the higher the entropy. A higher Kolmogorov entropy value means a higher rate of change of the internal structure and of the information content and thus the faster development of complexity. The following paragraph should help us to understand why Kolmogorov entropy is the rate of change of Shannon entropy of a system.

Based on the setting for Eq. 3.13, Shannon entropy for a fix n is

$$K_n = - \sum_{i_0, i_1, i_2, \dots, i_n} (p(i_0, i_1, i_2, \dots, i_n) \log p(i_0, i_1, i_2, \dots, i_n)) \quad (3.11)$$

Then, $K_{n+1} - K_n$ is the increment of Shannon entropy from time $n\tau$ to $(n+1)\tau$, which can be seen as the information needed to predict the status at $(n+1)\tau$ given the status at up to $n\tau$ is known. The overall rate of change H' of Shannon entropy is thus

$$H' = \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} \sum_{n=0}^{N-1} (K_{n+1} - K_n) \quad (3.12)$$

Consider

$$\begin{aligned} \sum_{n=0}^{N-1} (K_{n+1} - K_n) &= (K_1 - K_0) + (K_2 - K_1) + (K_3 - K_2) + \dots + (K_N - K_{N-1}) \\ &= K_N - K_0 \end{aligned}$$

and $K_0 = \sum_{i_0} (p(i_0) \log p(i_0)) = 0$ because status at $t = 0$ is always known and $p(i_0) = 1$, we have

$$\begin{aligned}
H' &= \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} \sum_{n=0}^{N-1} (K_{n+1} - K_n) = \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} (K_N - K_0) = \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} K_N \\
&= - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} \left(\sum_{i_0, i_1, i_2, \dots, i_N} (p(i_0, i_1, i_2, \dots, i_N) \log p(i_0, i_1, i_2, \dots, i_N)) \right) = K
\end{aligned}$$

That is, $K = H'$, which shows that Kolmogorov entropy is the rate of change of Shannon entropy of a system.

The application of Kolmogorov entropy in time series commonly goes through another quantity R_2 , which is a special case of the rate of change of Renyi entropy R_q .

For the setting for Eq. 3.9, the rate of change of Renyi entropy is

$$R_q = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N\tau} \frac{1}{q-1} \log \sum_{i_0, i_1, i_2, \dots, i_N} p^q(i_0, i_1, i_2, \dots, i_N) \quad (3.13)$$

From Eqs. 3.7 and 3.8, we have that R_1 is the Kolmogorov entropy and R_2 is the lower bound of Kolmogorov entropy.

Grassberger and Procaccia [6] demonstrated that for typical cases, R_2 is numerically close to K . More importantly, R_2 can be extracted fairly easily from an experimental signal as follows.

Let $X_i = (y_i, y_{i+1}, \dots, y_{i+d-1})$ (where i is from 1 to $N-d+1$) be a sequence of Y starting at y_i with a length of d . That is, we have a sequence of $N-d+1$ vectors, $X_1, X_2, \dots, X_{N-d+1}$.

Consider

$$C_d(\epsilon) = \lim_{N \rightarrow \infty} \left\{ \frac{1}{N^2} \times \text{number of pairs}(n, m) \text{ with } \left(\sum_{i=1}^d (X_{n+i} - X_{m+i})^2 \right)^{1/2} < \epsilon \right\}$$

$$K_{2,d}(\epsilon) = \frac{1}{\tau} \log \frac{C_d(\epsilon)}{C_{d+1}(\epsilon)}$$

Grassberger and Procaccia [6] proved that

$$\lim_{\substack{d \rightarrow \infty \\ \epsilon \rightarrow 0}} K_{2,d}(\epsilon) = K_2, \text{ i.e., } \lim_{\substack{d \rightarrow \infty \\ \epsilon \rightarrow 0}} \frac{1}{\tau} \log \frac{C_d(\epsilon)}{C_{d+1}(\epsilon)} = K_2 \quad (3.14)$$

The Euclidean distance in $C_d(\epsilon)$ may be replaced by the maximum norm [13].

3.6 Approximate Entropy

Consider a time series of data with length of N , $Y = (y_1, y_2, \dots, y_N)$, from measurements equally spaced in time. Let $X_i = (y_i, y_{i+1}, \dots, y_{i+m-1})$ (where i is from 1 to $N-m+1$) be a sequence of Y starting at y_i with a length of m . That is, we have a sequence of $N-m+1$ vectors, $X_1, X_2, \dots, X_{N-m+1}$. For each pair of sequences, X_i and X_j , we may define a distance between them. One commonly used distance is the maximum absolute difference of their corresponding elements, namely,

$$d(X_i, X_j) = \max(|X_i - X_j|) = \max_{k=1,2,\dots,m} |y_{i+k-1} - y_{j+k-1}|$$

For a sequence X_i , a sequence X_j that has a distance from X_i less than or equal to r (i.e., $d(X_i, X_j) \leq r$) is defined as within the r of X_i and count it as a match with X_i . We may count the number of X_j that matches with X_i with respect to (w.r.t.) r and denote it as $C_i^m(r)$. The proportion of X_j matching with X_i w.r.t. r is then

$$P_i^m(r) = C_i^m(r)/(N - m + 1)$$

When N is large, $P_i^m(r)$ represents the probability that any X_j matching with X_i w.r.t. r .

The average proportion of matches for all sequences X_i ($1 \leq i \leq N - m + 1$) is thus

$$P^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} P_i^m(r)$$

Define $\Phi^m(r)$ as

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

Then, Eckmann-Ruelle (ER) entropy [4] is

$$\begin{aligned} \text{ER entropy} &= \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} (\Phi^m(r) - \Phi^{m+1}(r)) \\ &= \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} \left(\frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log P_i^m(r) - \frac{1}{N - m} \sum_{i=1}^{N-m} \log P_i^{m+1}(r) \right) \end{aligned}$$

The approximate entropy for fixed m and r [9] is

$$\begin{aligned} \text{ApEn}(m, r) &= \lim_{N \rightarrow \infty} (\Phi^m(r) - \Phi^{m+1}(r)) \\ &= \lim_{N \rightarrow \infty} \left(\frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log P_i^m(r) - \frac{1}{N-m} \sum_{i=1}^{N-m} \log P_i^{m+1}(r) \right) \end{aligned}$$

Given N data points

$$\begin{aligned} \text{ApEn}(m, r, N) &= \Phi^m(r) - \Phi^{m+1}(r) \\ &= \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log P_i^m(r) - \frac{1}{N-m} \sum_{i=1}^{N-m} \log P_i^{m+1}(r) \end{aligned}$$

The theoretical framework for approximate entropy is based on the following theorems [9].

Assume a stationary process $u(i)$ with continuous state space. Let (X, Y) be the joint stationary probability measure on a two-dimensional space for this process (assuming uniqueness), and π_X be the equilibrium probability of X . Then a.s.

Theorem 1

$$\text{ApEn}(1, r) = - \int u(x, y) \log \left(\frac{\int_{z=y-r}^{y+r} \int_{w=x-r}^{x+r} u(w, z) dw dz}{\int_{w=x-r}^{x+r} \pi(w) dw} \right) dx dy$$

Theorem 2 For an *i.i.d.* process with density function, a.s. (for any $m \geq 1$)

$$\text{ApEn}(m, r) = - \int \pi(y) \log \left(\int_{z=y-r}^{y+r} \pi(z) dz \right) dy$$

Theorem 3 In the first-order stationary Markov chain (discrete state space values) case, with $r < \min(|x-y|, x \neq y, x$ and y state space values X), a.s. for any m

$$\text{ApEn}(m, r) = - \sum_{x \in X} \sum_{y \in Y} \pi(x) p_{xy} \log p_{xy}$$

Pincus [9] considers $\text{ApEn}(m, r)$ as a family of formulas and $\text{ApEn}(m, r, N)$ as a family of statistics; system comparisons are intended with fixed m and r . This family of statistics is rooted in the work of Grassberger and Procaccia (1983) [6, 7] and Eckmann and Ruelle [4]. The above theory and method for a measure of regularity proposed by Pincus [9] are closely related to the Kolmogorov entropy, the rate of generation of new information, which can be applied to the typically short and noisy time series of clinical data [11].

3.7 Sample Entropy

Like approximate entropy, sample entropy is also defined on the count $C_i^m(r)$ and proportion $P_i^m(r)$ of a sequence X_j matching with another sequence X_i w.r.t. r but with two alterations [11]. First, self-matches are counted in the calculation of $C_i^m(r)$ for approximate entropy, whereas self-matches are excluded in the calculation of $C_i^m(r)$ for sample entropy. Second, for a fixed m , only the first $N - m$ sequences X_j of length m are used for calculating $C_i^m(r)$ to ensure that the number of sequences used for $C_i^m(r)$ is the same as the number of sequences available for $C_i^{m+1}(r)$. To avoid confusion, in the altered setting as contrasted to the unaltered sitting, we may use $C_i^m(r)^*$ to denote the number of matches with the i^{th} sequence X_i of length m , i.e., $C_i^m(r)^* = \{C_i^m(r); i = 1, 2, i - 1, i + 1, N - m\}$, and to use $C_i^m(r)^{**}$ to denote the number of matches with the i^{th} sequence X_i sequence of length $m + 1$, i.e., $C_i^m(r)^{**} = \{C_i^{m+1}(r); i = 1, 2, i - 1, i + 1, N - m\}$. Similar to the proportion $P_i^m(r)$ of matches for approximate entropy, we have the proportions, $P_i^m(r)^*$ and $P_i^m(r)^{**}$, of matches for sample entropy as follows:

$$P_i^m(r)^* = C_i^m(r)^*/(N - m)$$

$$P_i^m(r)^{**} = C_i^m(r)^{**}/(N - m - 1)$$

Similar to the average proportion $P^m(r)$ of matches for all sequences X_i for approximate entropy, we have the average proportions, $P^m(r)^*$ and $P^m(r)^{**}$, of matches for sample entropy as follows:

$$P^m(r)^* = \frac{1}{N - m} \sum_{i=1}^{N-m} P_i^m(r)^*$$

$$P^m(r)^{**} = \frac{1}{N - m - 1} \sum_{i=1}^{N-m-1} P_i^m(r)^{**}$$

Clearly, $P^m(r)^*$ is an estimate for the probability that two different sequences will match for m points, whereas $P^m(r)^{**}$ is an estimate for the probability that two different sequences will match for $m + 1$ points.

The sample entropy for fixed m and r [11] is defined as

$$\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left(-\log \frac{P^m(r)^{**}}{P^m(r)^*} \right)$$

Given N data points

$$\text{SampEn}(m, r, N) = -\log \frac{P^m(r)^{**}}{P^m(r)^*}$$

From the above definition, sample entropy is essentially the negative logarithm of the ratio of the probability that any two different sequences in a time series match for m points to the probability that any two different sequences in the time series match for $m + 1$ points.

The theoretical framework for sample entropy is based on Grassberger and Procaccia's (1983) work as concluded in Eq. 3.14.

3.8 Multiscale Entropy Analysis

Kolmogorov entropy, approximate entropy, and sample entropy described above are based on a single scale, reflecting the uncertainty of the next point given the past history of the series as their calculation depends on a function's one step difference (e.g., $K_{n+1} - K_n$). Hence, they do not account for features related to structure on scales other than the shortest one [2]. To address this issue, Zhang [14] introduced multiscale entropy to the analysis of the physical time series. Costa et al. [2] further extended the multiscale entropy technique applicable to the analysis of the biological time series.

The basic idea for calculating multiscale entropy is to first generate a new time series consisting of the means of consecutive nonoverlap segments each with fixed length k of observed data points and then to calculate Kolmogorov entropy, approximate entropy, or sample entropy based on the newly generated time series. Specifically, we first divide the original signal ($\{y_i\}, 1 \leq i \leq N$) into nonoverlapping segments of equal length (k) and calculating the mean value of the data points in each of these segments [14]. This process is called coarse graining, and the newly generated time series is called coarse-grained time series. The length k is called a scale factor (or simply a scale). The coarse-graining process is repeated for a range of the scale factor. As the scale factor k changes, we will construct different coarse-grained time series, and subsequently we will calculate corresponding entropy values on the newly coarse-grained time series. We may now plot the entropy of coarse-grained time series against the scale factor k . This procedure is called multiscale analysis (MSE) [2].

The coarse-graining process can be applied to not only the mean of a divided segment but also its variance and any moments in statistics [3]. Costa and Goldberger [3] termed the MSE on mean-based coarse-grained time series as MSE_μ , MSE on variance-based coarse-grained time series as MSE_σ^2 , and MSE on any moment-based coarse-grained time series as MSE_n . Note, in statistics, the mean is the first central moment, and the variance is the second central moment.

3.9 Demonstration of an Example

Diabetes is a major disease in human being. It is a group of metabolic diseases in which the patients' blood sugar level is high over a prolonged period. The key index for monitoring diabetes is blood glucose level in a person. The left panel of Fig. 3.1 shows the glucose levels of a healthy person (black points) and a patient with type II diabetes (gray points), measured once every 3 min in 2 days. I calculate the multiscale sample entropy based on the data shown in the left panel of Fig. 3.1 and obtain the results shown in the right panel of Fig. 3.1.

The average value and standard deviation of the glucose levels are 5.16 and 1.00, respectively, in the healthy person and insulin and 12.45 and 3.08, respectively, in the patient with type II diabetes. The diabetes patient has sample entropy values lower than the healthy person at each scale (the right panel of Fig. 3.1). The change of direction in entropy is different from that in the average value (or variability), which may indicate that entropy contains information that the average value cannot be disclosed.

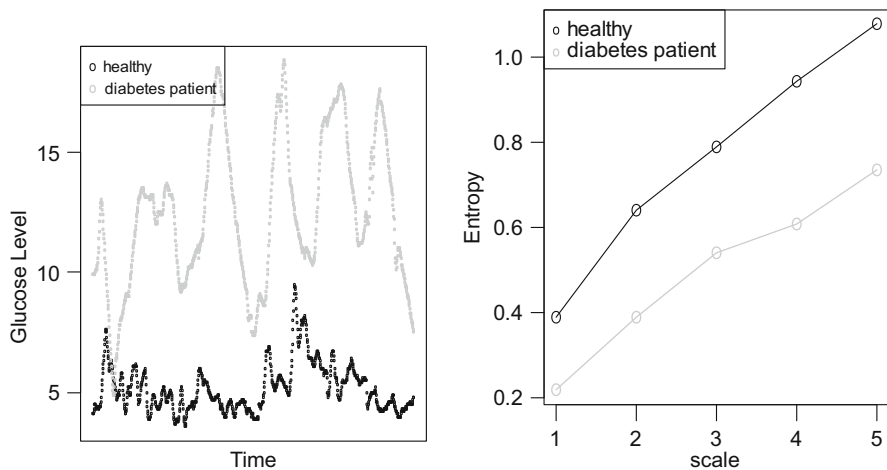


Fig. 3.1 Human glucose dynamics and its analysis using multiscale sample entropy. *Left Panel*, value of glucose levels in a healthy person (*black points*) and in a patient with type II diabetes (*gray points*) measured once in every 3 min in 2 days. *Right Panel*, multiscale sample entropy values in a healthy person (*black points and line*) and in a patient with type II diabetes (*gray points*) calculated based on the glucose levels shown in the *left panel*

3.10 Discussion and Conclusion

Based on the description of various concepts of entropy for biological dynamics, the Shannon entropy is a measurement of the disorder (or uncertainty) of a system. Shannon entropy is a special case (with $q = 1$) of more broad entropy called Renyi entropy. Kolmogorov entropy is the rate of change of Shannon entropy. The low bound of Kolmogorov entropy is the rate of change of Renyi entropy with $q = 2$. This low bound can be approximated by an entropy that can be calculated using time series data from biological dynamics. This entropy is called approximate entropy. Sample entropy is an improved version of approximate entropy that corrects unpleased self-matching effect. Both sample entropy and approximate entropy can be calculated in different layers of data, resulting in multiscale entropy (MSE). Sample entropy, approximate entropy, and/or their corresponding MSE have now been broadly used to assess the complexity of biological dynamics in various diseases measured by noninvasive medical device.

As demonstrated in the diabetes example, entropy analysis for the dynamics of physiological signals can disclose information that is not contained in the average value or variability of the physiological signals. Thus, entropy analysis could be used to differentiate the major diseases such as diabetes into different sub-diseases on the top of existing approach (such as using the average value in the diabetes case). It could also be used to reclassify the major diseases so that more specific and more effective drugs or treatment can be developed. All these will help to the development of precision medicine in which the right drug at the right dosage can be prescribed to treat the right person at the right time.

References

1. Clausius R (1854) Poggendorff's Ann., December 1854 xciii:481
2. Costa M, Goldberger AL, Peng CK (2002) Multiscale entropy analysis of complex physiologic time series. *Phys Rev Lett* 89(068):102
3. Costa M, Goldberger AL (2015) Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series. *Entropy* 17:1197–1203
4. Eckmann JP, Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Rev Mod Phys* 57:617–654
5. Elenko E, Underwood L, Zohar D (2015) Defining digital medicine. *Nat Biotechnol* 33:456–461
6. Grassberger P, Procaccia I (1983) Estimation of the Kolmogorov entropy from a chaotic signal. *Phys Rev A* 28:2591–2593
7. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. *Physica D* 9:189–208
8. Kolmogorov AN (1958) New metric invariant of transitive dynamical systems and endomorphisms of Lebesgue spaces. *Dokl Russ Acad Sci* 119(N5):861–864
9. Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci U S A* 88:2297–2301

10. Rényi A (1961) On measures of information and entropy. Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960. pp 547–561
11. Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 278:H2039–H2049
12. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
13. Takens F (1983) Invariants related to dimension and entropy. Proceedings of the Thirteenth Coloquio Brasileiro de Matematicas (Rio de Janerio)
14. Zhang YC (1991) Complexity and $1/f$ noise. A phase space approach. *J Phys I France* 1:971–977