

Information Distances versus Entropy Metric

Bo Hu ¹, Lvqing Bi ^{2,3} and Songsong Dai ^{3,*}

¹ School of Mechanical and Electrical Engineering, Guizhou Normal University, Guiyang 550025, China; hb.gznu@gmail.com

² School of Electronics and Communication Engineering, Yulin Normal University, Yulin 537000, China; bilvqing108@163.com

³ School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

* Correspondence: ssdai@stu.xmu.edu.cn; Tel.: +86-592-258-0135

Academic Editor: Raúl Alcaraz Martínez

Received: 18 April 2017; Accepted: 2 June 2017; Published: 7 June 2017

Abstract: Information distance has become an important tool in a wide variety of applications. Various types of information distance have been made over the years. These information distance measures are different from entropy metric, as the former is based on Kolmogorov complexity and the latter on Shannon entropy. However, for any computable probability distributions, up to a constant, the expected value of Kolmogorov complexity equals the Shannon entropy. We study the similar relationship between entropy and information distance. We also study the relationship between entropy and the normalized versions of information distances.

Keywords: entropy; Kolmogorov complexity; information distance; normalized information distance

1. Introduction

Information distance [1] is a universal distance measure between individual objects based on Kolmogorov complexity. It is the length of a shortest program that transforms one object into the other object. Since the theory of information distance was proposed, various information distance measures have been known. The normalized versions of information distances [2] have been introduced for measuring similarity between sequences. The min distance and its normalized version, which do not satisfy the triangle inequality, have been presented in [3,4]. The time-bounded version of information distance [5] has been used for studying the computability properties of the normalized information distances. A safe approximability of the normalized information distance have been discussed in [6]. Since the normalized information distance is uncomputable, two practical distance measures, the normalized compression distance and the Google similarity distance, have been presented [7–11]. These distance measures have been successfully applied to bioinformatics [10], music clustering [7–9], linguistics [2,12], plagiarism detection [13], question answering [3,4,14] and many more.

As mentioned in [1], information distance should be contrasted with the entropy metric. The former is based on Kolmogorov complexity and the latter on Shannon entropy. Various relations between Shannon entropy and Kolmogorov complexity are known [15–17]. It is well known that for any computable probability distributions the expected value of Kolmogorov complexity equals the Shannon entropy [18,19]. Linear inequalities are valid for Shannon entropy are also valid for Kolmogorov complexity, and vice versa [20]. We also know that various notions are both based on Shannon entropy and Kolmogorov complexity. Hence, many similar relationships between entropy based notions and Kolmogorov complexity based notions have been proposed. Relations between time-bounded entropy measures and time-bounded Kolmogorov complexity have been proposed in [21]. Relations between Shannon mutual information and algorithmic (Kolmogorov) mutual information have been proposed in [18]. Then, relations between entropy based cryptographic security and Kolmogorov complexity

based cryptographic security have been studied [22–25]. One-way functions have been studied on both time-bounded entropy and time-bounded Kolmogorov complexity [26]. However, the relationship between information distance and entropy has not been studied. In this paper, we study the similar relationship between information distance and the entropy metric. We also analyze the validity of the relationship between normalized information distance and the entropy metric.

The rest of this paper is organized as follows: In Section 2, some basic notions are reviewed. In Section 3, we study the relationship between information distance and the entropy metric. In Section 4, we study the relationship between normalized information distance and the entropy metric. Finally, conclusions are stated in Section 5.

2. Preliminaries

In this paper, let $|x|$ be the length of the string x and $\log(\cdot)$ be the function $\log_2(\cdot)$.

2.1. Kolmogorov Complexity

Kolmogorov complexity was introduced independently by Solomonoff [27] and Kolmogorov [28] and later by Chaitin [29]. Some basic notions of Kolmogorov complexity are given below. For more details, see [16,17]. We use the prefix-free definition of Kolmogorov complexity. A string x is a proper prefix of a string y if we have $y = xz$ for $z \neq \varepsilon$, where ε is the empty string. A set of strings A is prefix-free if there are not two strings x and y in A such that x is a proper prefix of y . For convenience, we use the prefix-free Turing machine, i.e., Turing machines with a prefix-free domain.

Let F be a fixed prefix-free optimal universal Turing machine. The conditional Kolmogorov complexity $K(y|x)$ of y given x is defined by

$$K(y|x) = \min\{|p| : F(p, x) = y\},$$

where $F(p, x)$ is the output of the program p with auxiliary input x when it is run in the machine F .

The (unconditional) Kolmogorov complexity $K(y)$ of y is defined as $K(y|\varepsilon)$.

2.2. Shannon Entropy

Shannon entropy [30] is a measure of the average uncertainty in a random variable. Some basic notions of entropy are given here. For more details, see [16,18]. For simplicity, all random variables mentioned in the paper are outcomes in the sets of finite strings.

Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, the marginal distributions of X and Y are defined by $f_1(x) = \sum_y f(x, y)$ and $f_2(x) = \sum_x f(x, y)$, respectively.

The joint Shannon entropy of X and Y is defined as

$$H(X, Y) = - \sum_{x \in X, y \in Y} f(x, y) \log f(x, y). \quad (1)$$

The Shannon entropy of X is defined as

$$H(X) = - \sum_{x \in X} f_1(x) \log f_1(x) = - \sum_{x \in X, y \in Y} f(x, y) \log f_1(x). \quad (2)$$

The conditional Shannon entropy with respect to Y given X is defined as

$$H(Y|X) = \sum_{x \in X} f_1(x) H(Y|X = x) \quad (3)$$

$$= - \sum_{x \in X} f_1(x) \sum_{y \in Y} f(y|x) \log f(y|x) \quad (4)$$

$$= - \sum_{x \in X, y \in Y} f(x, y) \log f(x|y). \quad (5)$$

The mutual information between variables X and Y is defined as

$$I(X; Y) = H(X) - H(X|Y). \quad (6)$$

Kolmogorov complexity and Shannon entropy are fundamentally different measures. However, for any computable probability distributions, up to $K(f) + O(1)$, the Shannon entropy equals the expected value of the Kolmogorov complexity [17–19]. Conditional Kolmogorov complexity and conditional Shannon entropy are also related.

The following two Lemmas are Theorem 8.1.1 from [17] and Theorem 5 from [22], respectively.

Lemma 1. Let X be a random variable over \mathcal{X} . For any computable probability distribution $f(x)$ over \mathcal{X} ,

$$0 \leq \left(\sum_x f(x) K(x) - H(X) \right) \leq K(f) + O(1). \quad (7)$$

Lemma 2. Let X, Y be two random variables over \mathcal{X}, \mathcal{Y} , respectively. For any computable probability distribution $f(x, y)$ over $\mathcal{X} \times \mathcal{Y}$,

$$0 \leq \left(\sum_{x,y} f(x, y) K(x|y) - H(X|Y) \right) \leq K(f) + O(1). \quad (8)$$

The following two Lemmas will be used in the next section.

Lemma 3. There are four positive integer a, b, c, d such that

$$\frac{1}{2} \max(a, b) + \frac{1}{2} \max(c, d) > \max\left(\frac{a+c}{2}, \frac{b+d}{2}\right).$$

Proof. Let $a > b > 0$ and $d > c > 0$, then $\frac{1}{2} \max(a, b) + \frac{1}{2} \max(c, d) = \frac{a+d}{2} > \max\left(\frac{a+c}{2}, \frac{b+d}{2}\right)$. \square

Lemma 4. There are four positive integer a, b, c, d such that

$$\frac{1}{2} \min(a, b) + \frac{1}{2} \min(c, d) < \min\left(\frac{a+c}{2}, \frac{b+d}{2}\right).$$

Proof. Let $a > b > 0$ and $d > c > 0$, then $\frac{1}{2} \min(a, b) + \frac{1}{2} \min(c, d) = \frac{b+c}{2} < \min\left(\frac{a+c}{2}, \frac{b+d}{2}\right)$. \square

3. Information Distance Versus Entropy

A metric on a set \mathcal{X} is a function $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ having the following properties: for every $x, y, z \in \mathcal{X}$

- (i) $d(x, y) = 0$ if and only if $x = y$;
- (ii) $d(x, y) = d(y, x)$;
- (iii) $d(x, y) + d(y, z) \geq d(x, z)$.

Here, entropy metric means the metric on the set of all random variables over a set. $d(X, Y) = H(X|Y) + H(Y|X)$ is a metric [16]. It is easy to know that $d(X, Y) = \max\{H(X|Y), H(Y|X)\}$ is also a metric.

Information distance $E_{\max}(x, y)$ [1], the length of a shortest program computing y from x and vice versa, is defined as

$$E_{\max}(x, y) = \min\{|p| : F(p, x) = y, F(p, y) = x\}.$$

In [1], up to an additive logarithmic term, the equality, $E_{\max}(x, y) = \max(K(x|y), K(y|x))$, holds. So $E_{\max}(x, y)$ is called the max distance between x and y .

We show the following relationship between max distance and the entropy metric.

Theorem 1. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\max(H(X|Y), H(Y|X)) \leq \sum_{x,y} f(x, y) E_{\max}(x, y) \leq H(X|Y) + H(Y|X) + 2K(f) + O(1)$$

Proof. First, from Lemma 2, we have

$$\sum_{x,y} f(x, y) E_{\max}(x, y) \geq \sum_{x,y} f(x, y) K(x|y) \geq H(X|Y), \quad (9)$$

$$\sum_{x,y} f(x, y) E_{\max}(x, y) \geq \sum_{x,y} f(x, y) K(y|x) \geq H(Y|X). \quad (10)$$

Thus

$$\sum_{x,y} f(x, y) E_{\max}(x, y) \geq \max(H(X|Y), H(Y|X)). \quad (11)$$

Moreover, from Lemma 2, we get

$$\begin{aligned} \sum_{x,y} f(x, y) E_{\max}(x, y) &\leq \sum_{x,y} f(x, y) (K(x|y) + K(y|x)) \\ &\leq \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x) \\ &\leq H(X|Y) + H(Y|X) + 2K(f) + O(1). \end{aligned}$$

□

Remark 1. From the above Theorem, the inequality $\sum_{x,y} f(x, y) E_{\max}(x, y) \geq \max(H(X|Y), H(Y|X))$ holds.

Unfortunately, the inequality $\sum_{x,y} f(x, y) E_{\max}(x, y) \leq \max(H(X|Y), H(Y|X))$ does not hold.

For instance, let the joint probability distribution $f(x, y)$ of X and Y be $f(x_1, y_1) = f(x_2, y_2) = 0.5$, and let $a = K(x_1|y_1)$, $b = K(y_1|x_1)$, $c = K(x_2|y_2)$ and $d = K(y_2|x_2)$ such that $a \neq b$ and $d \neq c$. Assume, without loss of generality, that $a > b$ and $d > c$, then from Lemma 3, we have $\sum_{x,y} f(x, y) E_{\max}(x, y) > \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x)$.

This means we will get the inequality $\sum_{x,y} f(x, y) E_{\max}(x, y) > \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x) \geq \max(H(X|Y), H(Y|X))$ for some cases.

From above results, we know that the relationship $\sum_{x,y} f(x, y) E_{\max}(x, y) \approx \max(H(X|Y), H(Y|X))$ does not hold.

Because the mutual information between X and Y is defined as $I(X; Y) = H(X) - H(X|Y)$, we have the following result.

Corollary 1. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\max(H(X), H(Y)) - I(X; Y) \leq \sum_{x,y} f(x, y) E_{\max}(x, y) \leq H(X) + H(Y) - 2I(X; Y) + 2K(u) + O(1)$$

Min distance $E_{\min}(x, y)$ [3,4] is defined as

$$E_{\min}(x, y) = \min\{|p| : F(p, x, z) = y, F(p, y, r) = x, |p| + |z| + |r| \leq E_{\max}(x, y)\}.$$

In [3,4], the equality, $E_{\min}(x, y) = \min(K(x|y), K(y|x))$, holds, when a term $O(\log|x| + |y|)$ is omitted. Then we have the following relationship between min distance and the entropy metric.

Theorem 2. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \min(H(X|Y), H(Y|X)) + K(f) + O(1)$$

Proof. From Lemma 1, we have

$$\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \sum_{x,y} f(x, y) K(x|y) \leq H(X|Y) + K(f) + O(1), \quad (12)$$

$$\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \sum_{x,y} f(x, y) K(y|x) \leq H(Y|X) + K(f) + O(1). \quad (13)$$

Thus

$$\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \min(H(X|Y), H(Y|X)) + K(f) + O(1). \quad (14)$$

□

Remark 2. From the above Theorem, the inequality $\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \min(H(X|Y), H(Y|X))$ holds.

Unfortunately, from Lemma 4, we know that the inequality $\sum_{x,y} f(x, y) E_{\min}(x, y) \geq \min(H(X|Y), H(Y|X))$ does not hold.

Thus, the relationship $\sum_{x,y} f(x, y) E_{\min}(x, y) \approx \min(H(X|Y), H(Y|X))$ does not hold.

Corollary 2. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x, y) E_{\min}(x, y) \leq \min(H(X), H(Y)) - I(X; Y) + K(f) + O(1)$$

Sum distance $E_{\text{sum}}(x, y)$ [1] is defined as

$$E_{\text{sum}}(x, y) = K(x|y) + K(y|x).$$

Then we have the following relationship between sum distance and entropy measure.

Theorem 3. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$H(X|Y) + H(Y|X) \leq \sum_{x,y} f(x, y) E_{\text{sum}}(x, y) \leq H(X|Y) + H(Y|X) + 2K(f) + O(1)$$

Proof. First, we have

$$\begin{aligned} \sum_{x,y} f(x, y) E_{\text{sum}}(x, y) &= \sum_{x,y} f(x, y) (K(x|y) + K(y|x)) \\ &= \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x) \end{aligned}$$

Then, from Lemma 1, we can get

$$0 \leq \sum_{x,y} f(x, y) E_{\text{sum}}(x, y) - (H(X|Y) + H(Y|X)) \leq 2K(f) + O(1). \quad (15)$$

□

Corollary 3. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$H(X) + H(Y) - 2I(X; Y) \leq \sum_{x,y} f(x, y) E_{\text{sum}}(x, y) \leq H(X) + H(Y) - 2I(X; Y) + 2K(f) + O(1)$$

From above results we know that, when f is given, up to a additive constant,

$$\sum_{x,y} f(x, y) E_{\text{sum}}(x, y) = H(X|Y) + H(Y|X) = H(X) + H(Y) - 2I(X; Y). \quad (16)$$

4. Normalized Information Distance Versus Entropy

In this section, we establish relationships between entropy and the normalized versions of information distances.

The normalized version $e_{\text{max}}(x, y)$ [2] of $E_{\text{max}}(x, y)$ is defined as

$$e_{\text{max}}(x, y) = \frac{\max(K(x|y), K(y|x))}{\max(K(x), K(y))}.$$

Theorem 4. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x, y) e_{\text{max}}(x, y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{\max(H(X), H(Y))}$$

Proof. First, we have

$$e_{\text{max}}(x, y) = \frac{\max(K(x|y), K(y|x))}{\max(K(x), K(y))} \leq \frac{K(x|y) + K(y|x)}{K(x)}. \quad (17)$$

Then

$$K(x) e_{\text{max}}(x, y) \leq K(x|y) + K(y|x). \quad (18)$$

Therefore,

$$\begin{aligned} \sum_{x,y} f(x, y) [K(x) e_{\text{max}}(x, y)] &\leq \sum_{x,y} f(x, y) (K(x|y) + K(y|x)) \\ &= \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x), \end{aligned}$$

From Lemmas 1 and 2, we have

$$\begin{aligned} H(X) \left[\sum_{x,y} f(x, y) e_{\text{max}}(x, y) \right] &\leq \left[\sum_{x,y} f(x, y) K(x) \right] \left[\sum_{x,y} f(x, y) e_{\text{max}}(x, y) \right], \text{ by Lemma 1} \\ &\leq \sum_{x,y} f(x, y) [K(x) e_{\text{max}}(x, y)] \\ &\leq \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x) \\ &\leq H(X|Y) + H(Y|X) + 2K(f) + O(1), \text{ by Lemma 2.} \end{aligned}$$

Then

$$\sum_{x,y} f(x, y) e_{\text{max}}(x, y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{H(X)}. \quad (19)$$

Similarly, we have

$$\sum_{x,y} f(x,y) e_{\max}(x,y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{H(Y)}. \quad (20)$$

Thus

$$\sum_{x,y} f(x,y) e_{\max}(x,y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{\max(H(X), H(Y))}. \quad (21)$$

□

Corollary 4. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x,y) e_{\max}(x,y) \leq \frac{H(X) + H(Y) - 2I(X; Y) + 2K(f) + O(1)}{\max(H(X), H(Y))}$$

The normalized version $e_{\min}(x, y)$ [3,4] of $E_{\min}(x, y)$ is defined as

$$e_{\min}(x, y) = \frac{\min(K(x|y), K(y|x))}{\min(K(x), K(y))}.$$

Because $e_{\min}(x, y) \leq e_{\max}(x, y)$, for all x, y [3,4], the following Corollary is straightforward with the above Theorem.

Corollary 5. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x,y) e_{\min}(x,y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{\max(H(X), H(Y))}$$

The normalized version $e_{\text{sum}}(x, y)$ [2] of $E_{\text{sum}}(x, y)$ is defined as

$$e_{\text{sum}}(x, y) = \frac{K(x|y) + K(y|x)}{K(x, y)}.$$

Theorem 5. Let X, Y be two random variables with a computable joint probability distribution $u(x, y)$, then

$$\sum_{x,y} f(x,y) e_{\text{sum}}(x,y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{H(X, Y)} \quad (22)$$

Proof. Since $e_{\text{sum}}(x, y) = \frac{K(x|y) + K(y|x)}{K(x, y)}$, then

$$K(x, y) e_{\text{sum}}(x, y) = K(x|y) + K(y|x). \quad (23)$$

Therefore,

$$\sum_{x,y} f(x,y) [K(x, y) e_{\text{sum}}(x, y)] = \sum_{x,y} f(x,y) K(x|y) + \sum_{x,y} f(x,y) K(y|x). \quad (24)$$

From Lemma 1, we have

$$\begin{aligned}
 H(X, Y) [\sum_{x,y} f(x, y) e_{sum}(x, y)] &\leq [\sum_{x,y} f(x, y) K(x, y)] [\sum_{x,y} f(x, y) e_{sum}(x, y)] \\
 &\leq \sum_{x,y} f(x, y) [K(x, y) e_{sum}(x, y)] \\
 &= \sum_{x,y} f(x, y) K(x|y) + \sum_{x,y} f(x, y) K(y|x) \\
 &\leq H(X|Y) + H(Y|X) + 2K(f) + O(1)
 \end{aligned}$$

Thus

$$\sum_{x,y} f(x, y) e_{sum}(x, y) \leq \frac{H(X|Y) + H(Y|X) + 2K(f) + O(1)}{H(X, Y)}. \quad (25)$$

□

Corollary 6. Let X, Y be two random variables with a computable joint probability distribution $f(x, y)$, then

$$\sum_{x,y} f(x, y) e_{sum}(x, y) \leq \frac{H(X, Y) - I(X; Y) + 2K(f) + O(1)}{H(X, Y)}. \quad (26)$$

5. Conclusions

As we know, the Shannon entropy of a distribution is approximately equal to the expected Kolmogorov complexity, up to a constant term that only depends on the distribution [17]. We studied whether a similar relationship holds for information distance. Theorem 5 gave the analogous result for sum distance. We also gave some bounds for the expected value of other (normalized) information distances.

Acknowledgments: The authors would like to thank the referees for their valuable comments and recommendations. Project supported by the Guangxi University Science and Technology Research Project (Grant No. 2013YB193) and the Science and Technology Foundation of Guizhou Province, China (LKS [2013] 35).

Author Contributions: Conceptualization, formal analysis, investigation and writing the original draft is done by Songsong Dai; Validation, review and editing is done by Bo Hu and Lvqing Bi; Project administration and Resources are provided by Bo Hu and Lvqing Bi. All authors have read and approved the final manuscript.

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

References

1. Bennett, C.H.; Gacs, P.; Li, M.; Vitányi, P.M.B.; Zurek, W. Information distance. *IEEE Trans. Inf. Theory* **1998**, *44*, 1407–1423.
2. Li, M.; Chen, X.; Li, X.; Ma, B.; Vitányi, P.M.B. The similarity metric. *IEEE Trans. Inf. Theory* **2004**, *50*, 3250–3264.
3. Li, M. Information distance and its applications. *Int. J. Found. Comput. Sci.* **2011**, *18*, 1–9.
4. Zhang, X.; Hao, Y.; Zhu, X.Y.; Li, M. New information distance measure and its application in question answering system. *J. Comput. Sci. Technol.* **2008**, *23*, 557–572.
5. Terwijn, S.A.; Torenvliet, L.; Vitányi, P.M.B. Nonapproximability of the normalized information distance. *J. Comput. Syst. Sci.* **2011**, *77*, 738–742.
6. Bloem, P.; Mota, F.; Rooij, S.D.; Antunes, L.; Adriaans, P. A safe approximation for Kolmogorov complexity. In Proceedings of the International Conference on Algorithmic Learning Theory, Bled, Slovenia, 8–10 October 2014; Auer, P., Clark, A., Zeugmann, T., Zilles, S., Eds.; Springer: Cham, Switzerland, 2014; Volume 8776, pp. 336–350.
7. Cilibrasi, R.; Vitányi, P.M.B.; Wolf, R. Algorithmic clustering of music based on string compression. *Comput. Music J.* **2004**, *28*, 49–67.
8. Cuturi, M.; Vert, J.P. The context-tree kernel for strings. *Neural Netw.* **2005**, *18*, 1111–1123.

9. Cilibrasi, R.; Vitányi, P.M.B. Clustering by compression. *IEEE Trans. Inf. Theory* **2005**, *51*, 1523–1545.
10. Li, M.; Badger, J.; Chen, X.; Kwong, S.; Kearney, P.; Zhang, H. An information-based sequence distance and its application to whole mito chondrial genome phylogeny. *Bioinformatics* **2001**, *17*, 149–154.
11. Cilibrasi, R.L.; Vitányi, P.M.B. The Google similarity distance. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 370–383.
12. Benedetto, D.; Caglioti, E.; Loreto, V. Language trees and zipping. *Phys. Rev. Lett.* **2002**, *88*, 048702.
13. Chen, X.; Francia, B.; Li, M.; McKinnon, B.; Seker, A. Shared information and program plagiarism detection. *IEEE Trans. Inf. Theory* **2004**, *50*, 1545–1551.
14. Bu, F.; Zhu, X.Y.; Li, M. A new multiword expression metric and its applications. *J. Comput. Sci. Technol.* **2011**, *26*, 3–13.
15. Leung-Yan-Cheong, S.K.; Cover, T.M. Some equivalences between Shannon entropy and Kolmogorov complexity. *IEEE Trans. Inf. Theory* **1978**, *24*, 331–339.
16. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: Hoboken, NJ, USA, 2006.
17. Li, M.; Vitányi, P.M.B. *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed.; Springer: New York, NY, USA, 2008.
18. Grünwald, P.; Vitányi, P. Shannon information and Kolmogorov complexity. *arXiv* **2008**, arXiv:cs/0410002.
19. Teixeira, A.; Matos, A.; Souto, A.; Antunes, L. Entropy measures vs. Kolmogorov complexity. *Entropy* **2011**, *13*, 595–611.
20. Hammer, D.; Romashchenko, A.; Shen, A.; Vereshchagin, N. Inequalities for Shannon entropies and Kolmogorov complexities. *J. Comput. Syst. Sci.* **2000**, *60*, 442–464.
21. Pinto, A. Comparing notions of computational entropy. *Theory Comput. Syst.* **2009**, *45*, 944–962.
22. Antunes, L.; Laplante, S.; Pinto, A.; Salvador, L. Cryptographic Security of Individual Instances. In *Information Theoretic Security*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 195–210.
23. Kaced, T. Almost-perfect secret sharing. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings (ISIT), Saint Petersburg, Russia, 31 July–5 August 2011; pp. 1603–1607.
24. Dai, S.; Guo, D. Comparing security notions of secret sharing schemes. *Entropy* **2015**, *17*, 1135–1145.
25. Bi, L.; Dai, S.; Hu, B. Normalized unconditional ϵ -security of private-key encryption. *Entropy* **2017**, *19*, 100.
26. Antunes, L.; Matos, A.; Pinto, A.; Souto, A.; Teixeira, A. One-way functions using algorithmic and classical information theories. *Theory Comput. Syst.* **2013**, *52*, 162–178.
27. Solomonoff, R. A formal theory of inductive inference—Part I. *Inf. Control* **1964**, *7*, 1–22.
28. Kolmogorov, A. Three approaches to the quantitative definition of information. *Probl. Inf. Transm.* **1965**, *1*, 1–7.
29. Chaitin, G. On the length of programs for computing finite binary sequences: Statistical considerations. *J. ACM* **1969**, *16*, 145–159.
30. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656.

