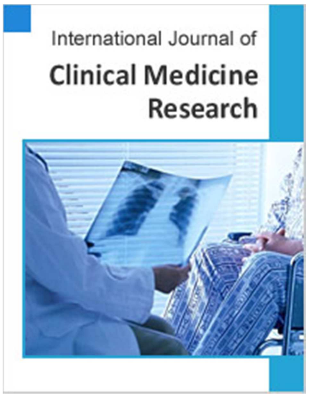
**International Journal of Clinical Medicine Research**
2017; 4(6): 72-75
http://www.aascit.org/journal/ijcmr
ISSN: 2375-3838

American Association for Science and Technology




**Keywords**

Schizophrenia,
Asthma,
Information Dimension,
Entropy,
Fractals,
Complex Networks

Received: September 25, 2017
Accepted: November 3, 2017
Published: November 16, 2017

# Erroneous Definition of the Information Dimension in Two Medical Applications

**Eric Rosenberg**

AT&T Labs, Middletown, NJ, USA

## Email address

ericr@att.com

## Citation

Eric Rosenberg. Erroneous Definition of the Information Dimension in Two Medical Applications. *International Journal of Clinical Medicine Research.* Vol. 4, No. 6, 2017, pp. 72-75.

## Abstract

The information dimension $d_I$ of a geometric object (e.g., grey matter) is one of several fractal dimensions that have been used in medicine. The information dimension $d_I$ is computed from N data points (e.g., pixels or voxels) by imposing a uniform grid (with grid box size s) covering the N points, measuring the number of points in each box of the grid, computing the resulting probability distribution, computing the entropy H(s) associated with this probability distribution and box size s, and then examining how H(s) scales with log s. The slope of the −H(s) vs. log s curve is the estimate of $d_I$. The purpose of this paper is to highlight two studies which used an incorrect definition of $d_I$: they defined $d_I$ to be the slope of the –log H(s) vs. log s curve. These studies are a 2016 study relating $d_I$ of subcortical grey matter structures to schizophrenia, and a 2005 study relating $d_I$ of human lungs to asthma. This paper first reviews the theory and application of computing $d_I$ for geometric objects, and then points out the error in these two studies. Given the attention now devoted to complex network models of the brain and other biological systems, recent results on computing $d_I$ and other fractal dimensions for a complex network are also briefly discussed.

## 1. Introduction

The information dimension $d_I$, originally defined by Balatoni and Renyi in 1956 and popularized by Farmer in 1982 ([4], [5]), has been an important tool for physicists ([1], [8], [9], [13], [18]). However, this fractal dimension has been applied far less frequently than the box counting dimension or the correlation dimension. Perhaps because it is less well known, two studies applying $d_I$ to medicine used an incorrect definition of this dimension: a 2005 study relating to asthma, and a 2016 study relating to schizophrenia. Their use of incorrect definitions of $d_I$, which apparently has not previously been recognized, means that other researchers attempting to duplicate their results will not be successful if these other researchers utilize the correct definition of $d_I$. Given the obvious importance of highlighting an error in the application of fractal dimensions to medicine, this paper first reviews the computation of $d_I$ for geometric objects and then points out the error in these 2005 and 2016 studies. Also, given the attention now devoted to complex network models of the brain (e.g., [12], [17]), recent results on computing $d_I$ and other fractal dimensions for a complex network are also briefly discussed.

## 2. The Information Dimension of a Geometric Object

As explained by Farmer [4], an information dimension $d_I$ can be computed for a probability distribution. Farmer describes how $d_I$ arises from the measurement process for a physical system, as follows. Suppose the region under study is contained in an $E$-dimensional "large" hypercube of side length L. For s < 1, the "large" hypercube is partitioned into "small" $E$-dimensional hypercubes, where the side length of each "small" hypercube is sL. Thus the partition contains $\left(\frac{L}{sL}\right)^{E} = s^{-E}$ "small" $E$-dimensional hypercubes, called "boxes". A probability is associated with each of the $s^{-E}$ boxes: let $p_j(s)$ be the probability associated with box $B_j$. Define $B(s) \equiv \{B_j \mid p_j(s) > 0\}$, so B(s) is the set of boxes with nonzero probability. The set $\{p_j(s) \mid B_j \in B(s)\}$ is known as the coarse-grained probability distribution at resolution s. The entropy H(s) of the probability distribution is defined by

$$H(s) \equiv -\sum_{B_j \in B(s)} p_j(s) \log p_j(s) \tag{1}$$

where the sum is over all boxes $B_j$ in B(s). The information dimension $d_I$ is defined by

$$d_I \equiv -\lim_{s \to 0} \frac{H(s)}{\log s} \tag{2}$$

assuming this limit exists. Thus $d_I$ is the rate at which the information scales as the precision of measurement is increased [4].

More insight into the fractal dimension $d_I$ arises from recalling that $H(s) = -E \log_2 s$ bits are needed to specify the position of a point in the unit hypercube in $R^E$ to an accuracy of s, where s < 1. The expression $H(s) = -E \log_2 s$ corresponds to the special case of (1) for which each of the $s^{-E}$ boxes with side length s in the $E$-dimensional unit hypercube is equally likely, with probability $p_j(s) = s^E$. For then the number |B(s)| of boxes is $s^{-E}$ and

$$H(s) = -\sum_{B_j \in B(s)} p_j(s) \log p_j(s) = -\sum_{B_j \in B(s)} s^E \log_2 s^E = -\log_2 s^E = -E \log_2 s.$$

Thus $d_I$ is the limit, as $s \to 0$, of the expected number of bits needed to specify the position of a point to accuracy s [20].

Given N points in $R^E$ from a geometric object (e.g., a human lung, or brain structure, or the retina), $d_I$ can be computed as follows. Cover the N points with a uniform grid of $E$-dimensional boxes with side length s. Discard any box containing none of the N points, and let B(s) denote the remaining set of nonempty boxes. For box $B_j \in B(s)$, define the probability $p_j(s)$ by $p_j(s) \equiv N_j(s)/N$, where $N_j(s)$ is the number of points contained in box $B_j$. Then compute the entropy H(s) using definition (1), and repeat this process for different box sizes. If a straight line is fitted to the plot of −H(s) vs. log s, over the range where this plot is roughly linear, the slope of the line is the estimate of $d_I$. When the distribution has a long tail, reflecting many very improbable boxes, this computation may be more efficient than box counting, since box counting requires counting each box containing at least one of the N points.

Definition (2) of $d_I$ can also be obtained from the generalized dimensions defined in 1983 by Grassberger [7] and also by Hentschel and Procaccia [11]. The generalized dimension $D_q$ is defined, for $q \neq 1$, by

$$D_q = \frac{1}{q-1} \lim_{s \to 0} \frac{\log\left[\sum_{B_j \in B(s)} p_j^q(s)\right]}{\log s}. \tag{3}$$

$D_1$ is evaluated by applying L'Hospital's rule, which yields

$$D_1 = \lim_{s \to 0} \frac{\sum_{B_j \in B(s)} p_j(s) \log p_j(s)}{\log s} = \lim_{s \to 0} \frac{-H(s)}{\log s}, \tag{4}$$

so it follows by (2) that $D_1 = d_I$. (Setting q = 0 in (3) yields

$$\sum_{B_j \in B(s)} p_j^q(s) = |B(s)|$$

where |B(s)| is the number of occupied boxes, so $D_0$ is the box counting dimension.)

## 3. Two Incorrect Definitions of the Information Dimension

This section discusses the two studies with an incorrect definition of $d_I$. In 2016, Zhao et al. [21] used $d_I$ to study schizophrenia. Three-dimensional MRIs were obtained from 38 people, 19 with schizophrenia, and 19 without. They computed $d_I$, using a 3-dimensional grid, for different sub-cortical areas of the brain, especially the hippocampus, which contributes to the encoding, consolidation, and retrieval of memory, and the representation of the temporal context. They chose to compute $d_I$, rather than some other fractal dimension, because $d_I$ has been shown to have a higher inter-class correlation [6]. To estimate $d_I$, points that deviate from the regression line were repeatedly excluded, and a new regression line computed, until most data points could be fitted. The results of their study included a finding that $d_I$ is significantly lower in the left and right hippocampus and in the left thalmus in patients with schizophrenia, compared with the healthy patients. However, they used an incorrect definition of $d_I$, namely

$$d_I = -\lim_{s \to 0} \frac{\log H(s)}{\log s}. \tag{5}$$

This incorrect definition is not a single "typo" or misprint, since "log H(s)" is used throughout [21].

A similar error in the computation of $d_I$ was made by Boser et al. [2], who in 2005 used $d_I$ to study differences between the lungs of three groups of deceased human non-smokers: fatal asthma (people who died from asthma), non-fatal asthma, and non-asthma control. From lung images of 1280 × 960 pixels, $d_I$ was calculated using 9 box sizes, ranging from about 10 pixels to about 100 pixels. They calculated $d_I$ = 1.83 for the non-asthma control group, $d_I$ = 1.76 for the non-fatal asthma group, and $d_I$ = 1.72 for the fatal asthma group. A statistical analysis showed that $d_I$ for the fatal asthma group and non-fatal asthma group are significantly lower than $d_I$ for the non-asthma group.

However, the difference in $d_I$ for the fatal asthma group and non-fatal asthma group is not significantly different. They observed that an advantage of using $d_I$ rather the box counting dimension is that $d_I$ is less likely to result in a false detection of multifractal characteristics. They also observed, as have many other researchers, that the exact interpretation of a fractal dimension, when used to quantify structure or change in structure, is less important than the ability of a fractal dimension to discriminate between states. However, their definition of $d_I$ is ([2], p. 818)

$$\log H(s) \approx -d_I \log s, \tag{6}$$

which is also incorrect, since the left hand side of (6) should be "H(s)" and not "log H(s)". This is also not a misprint: the graphs in [2] plot log H(s) vs. log s.

The fact that incorrect definitions of $d_I$ were used in [2] and [21] does not necessarily negate their findings. However, it does mean that their findings are valid only with respect to their definition of $d_I$, and other researchers may obtain very different conclusions when using the correct definition of $d_I$.

## 4. The Information Dimension of a Complex Network

This section briefly discusses the computation of $d_I$ for a complex network **G**, where a complex network is an arbitrary network without special structure (as opposed to, e.g., a regular lattice), for which all arcs have unit cost (so the length of a shortest path between two nodes is the number of arcs in that path), and all arcs are undirected (so the arc between nodes i and j can be traversed in either direction). The first step in computing $d_I$ is to determine, for a range of positive integer values of s, the minimal number of subnetworks (called "boxes") of diameter s needed to cover **G**. This step is called "box counting", and a wide variety of box counting methods are available [19]. Let N be the number of nodes in the network, and let $N_j(s)$ be the number of nodes in box $B_j$ of the minimal covering of **G** by boxes of diameter at most s-1. The probabilities $p_j(s)$ are defined by $p_j(s) = N_j(s)/N$, and H(s) is then computed using (1). The information dimension $d_I$ is the slope of the $-H(s)$ vs. log s plot, over the range of s for which this plot is roughly linear.

However, as shown in [14], there are in general multiple minimal coverings of **G**, and these different minimal coverings can yields different values of $d_I$. For a given integer box size s, a unique minimal covering of **G** can be obtained by computing the minimal covering that also maximizes the entropy H(s) [14], and computing this unique minimal covering can be accomplished by a simple modification of whatever box counting method is used. Unfortunately, even computing a maximal entropy minimal covering for each s does not eliminate ambiguity in the computation of $d_I$, since the choice of upper and lower bounds on the box sizes used to estimate the slope of the H(s) vs. log s curve can also lead to different values of $d_I$ [16]. Similar ambiguities arise in computing the generalized dimensions of **G** [15]. Thus, while the use of complex networks opens the door to computing a wide range of network characterizations [3], fractal dimensions of **G** must be computed and interpreted with the same care [10] required to compute and analyze fractal dimensions of geometric objects such as the brain or lungs.

## 5. Conclusion

The information dimension $d_I$ is a useful metric that can be calculated for a geometric object (e.g., a human lung, or brain structure, or the retina). The information dimension can also be calculated for complex networks, which are very useful for modelling biological systems. Unfortunately, incorrect definitions of $d_I$ were applied in a 2005 study of asthma, and in a 2016 study of schizophrenia. The error in these two studies, which apparently has not previously been recognized, means that other researchers attempting to duplicate their results will not be successful if these other researchers utilize the correct definition of $d_I$. This paper highlights these errors and provides the correct definition of the information dimension.

## Acknowledgements

Many thanks to Robert Murray, Kartik Pandit, and Curtis Provost for their comments and suggestions.

## References

[1] F. A. Bais and J. D. Farmer, "The Physics of Information", chapter in Philosophy of Information, edited by P. Adriaans and J. van Benthem (Elsevier, Oxford, United Kingdom, 2008).

[2] S. R. Boser, H. Park, S. F. Perry, M. C. Menache, and F. H. Y. Green, "Fractal Geometry of Airway Remodelling in Human Asthma", American Journal of Respiratory and Critical Care Medicine 172 (2005) pp. 817-823.

[3] L. da F. Costa, F. A. Rodrigues, G. Travieso and P. R. Villas Boas, "Characterization of Complex Networks: A Survey of Measurements", Advances in Physics 56 (2007) pp. 167-242.

[4] J. D. Farmer, "Dimension, Fractal Measures, and Chaotic Dynamics", in Evolution of Order and Chaos, Springer Series in Synergetics 17 (1982) pp. 228-246.

[5] J. D. Farmer, "Information Dimension and the Probabilistic Structure of Chaos", Z. Naturforsch. 37a (1982) pp. 1304-1325.

[6] J. Goni, O. Sporns, H. Cheng, M. Aznarez-Sanado, Y. Wang, S. Josa, G. Arrondo, V. P. Mathews, T. A. Hummer, W. G. Kronenberger, A. Avena-Koenigsberger, A. J. Saykin, and M. A. Pastor, "Robust Estimation of Fractal Measures for Characterizing the Structural Complexity of the Human Brain: Optimization and Reproducibility", NeuroImage 83 (2013) pp. 646-657.

[7] P. Grassberger, "Generalized Dimensions of Strange Attractors", Physics Letters 97A (1983) pp. 227-230.

[8] P. Grassberger and I. Procaccia, "Characterization of Strange Attractors", Physical Review Letters 50 (1983) pp. 346-349.

[9] P. Grassberger and I. Procaccia, “Measuring the Strangeness of Strange Attractors”, Physica 9D (1983) pp. 189-208.

[10] J. M. Halley, S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis, “Uses and Abuses of Fractal Methodology in Ecology", Ecology Letters 7 (2004) pp. 254-271.

[11] H. G. E. Hentschel and I. Procaccia, “The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors”, Physica D 8 (1983) pp. 435-444.

[12] E. A. Karuza, S. L. Thompson-Schill, and D. S. Bassett, “Local Patterns to Global Architectures: Influences of Network Topology on Human Learning", Trends in Cognitive Sciences 20 (2016) pp. 629-640.

[13] H. O. Peitgen, H. Jurgens, and D. Saupe, Chaos and Fractals (Springer-Verlag, New York, 1992).

[14] E. Rosenberg, “Maximal Entropy Coverings and the Information Dimension of a Complex Network”, Physics Letters A 381 (2017) pp. 574-580.

[15] E. Rosenberg, “Minimal Partition Coverings and Generalized Dimensions of a Complex Network”, Physics Letters A 381 (2017) pp. 1659-1664.

[16] E. Rosenberg, “Non-monotonicity of the Generalized Dimensions of a Complex Network”, Physics Letters A 381 (2017) pp. 2222-2229.

[17] M. Rubinov and O. Sporns, “Complex Network Measures of Brain Connectivity: Uses and Interpretations”, NeuroImage 52 (2010) pp. 1059-1069.

[18] D. Ruelle, “Deterministic Chaos: The Science and the Fiction (The 1989 Claude Bernard Lecture)”, Proc. R. Soc. Lond. A 427 (1990) pp. 241-248.

[19] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, “How to Calculate the Fractal Dimension of a Complex Network: the Box Covering Algorithm”, Journal of Statistical Mechanics: Theory and Experiment (2007) P03006.

[20] J. Theiler, “Estimating Fractal Dimension”, J. Optical Society of America A 7 (1990) pp. 1055-1073.

[21] G. Zhao, K. Denisova, P. Sehatpour, J. Long, W. Gui, J. Qiao, D. C. Javitt, and Z. Wang, “Fractal Dimension Analysis of Subcortical Gray Matter Structures in Schizophrenia”, PLOS ONE 11 (2016) e0155415.