Entropy, Fractals, & Extraterrestrial Life

D. Baron

Department of Mathematics
Western Washington University

May 16, 2016

Introduction

My project was inspired by the paper "The potential for detecting 'life as we don't know it' by fractal complexity analysis" (International Journal of Astrobiology, 2013) by Azua-Bustos and Vega-Martinez.

I've highlighted some interesting passages in that paper. . .

Outline

1. Mathematical Background
   - Shannon Entropy
   - Fractals
2. Fractal Image Analysis
   - The Method
   - Use in Astrobiology

Entropy of a Discrete Probability Space

In 1948, Claude Shannon introduced a quantity called "entropy", a measure of the uncertainty or surprise in an information source:

Definition. Let $\mathcal{X} = (X, p)$ be a discrete probability space; that is, X is the "alphabet" of possible outcomes, and for each x in X, the probability that we get x is $p(x)$.
The *entropy* of $\mathcal{X}$ is

$$H(\mathcal{X}) = -\sum_{x \in X} p(x) \log p(x).$$

Note that entropy has units, determined by the choice of logarithm base: bits, digits, nats, etc.

Examples

If there are n equally likely outcomes, then $p(x)$ is identically $1/n$ and we have

$$H(\mathcal{X}) = -\sum_{i=1}^{n} \frac{1}{n} \left(\log \frac{1}{n} \right) = \log n.$$

In fact, this is maximal: $|\mathcal{X}| = n \implies H(\mathcal{X}) \leq \log n$, with equality only when all outcomes are equiprobable.
The entropy of a fair coin toss is 1 bit; if the coin is biased, there is less entropy.

Examples

More interesting: if $X = \mathbb{N}$ and $p(n) = 2^{-n}$, then

$$H(\mathcal{X}) = \sum_{n=1}^{\infty} 2^{-n} \log 2^{n} = \sum_{n=1}^{\infty} 2^{-n} n \text{ bits} = 2 \text{ bits.}$$

Infinite "alphabet", finite—and quite small—uncertainty!

Properties of H

- H is nonnegative, and $H = 0$ only if the outcome is certain
- H is continuous in the $p(x)$
- H is additive: the entropy of m fair coin tosses is m bits.

It's these properties that make H useful as a measure of uncertainty.

Joint & Conditional Entropy

The *joint entropy* of spaces $\mathcal{X}$ and $\mathcal{Y}$ is

$$H(\mathcal{X}, \mathcal{Y}) = -\sum_{x,y} p(x,y) \log p(x,y),$$

and the *conditional entropy* of $\mathcal{Y}$ given $\mathcal{X}$ is

$$H(\mathcal{Y} \mid \mathcal{X}) = \mathbb{E}_{\mathcal{X}}[H(\mathcal{Y})] = -\sum_{x,y} p(x,y) \log p(y \mid x).$$

We then have $H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y} \mid \mathcal{X})$.

Entropy of a Message

Consider a sequence of probability spaces $\{\mathcal{X}_i\}_{i=1}^{\infty}$ with common "alphabet" X and respective probability functions p_i. Denote the kth joint space $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ by $\mathcal{Y}_k$. Then we have

$$\begin{aligned}
H(\mathcal{Y}_k) &= H(\mathcal{Y}_{k-1}) + H(\mathcal{X}_k \mid \mathcal{Y}_{k-1}) \\
&= H(\mathcal{Y}_{k-2}) + H(\mathcal{X}_{k-1} \mid \mathcal{Y}_{k-2}) + H(\mathcal{X}_k \mid \mathcal{Y}_{k-1}) \\
&\vdots \\
&= H(\mathcal{X}_1) + \sum_{i=2}^{k} H(\mathcal{X}_i \mid \mathcal{Y}_{i-1}).
\end{aligned}$$

Entropy of an Information Source

The infinite joint space $\mathcal{Y} = (\mathcal{X}_1, \ldots)$, the "information source", is then said to have per-symbol entropy

$$H(\mathcal{Y}) = \lim_{k\to\infty} \frac{H(\mathcal{Y}_k)}{k},$$

provided this limit exists.
In the special case when all allowed words of length k are equiprobable, we know that $H(\mathcal{Y}_k) = \log |Y_k|$, so the limit is

$$H(\mathcal{Y}) = \lim_{k\to\infty} \frac{\log |Y_k|}{k}.$$

Fun fact: this result holds for a *much* larger class of information sources. Read Shannon's paper if you're curious!

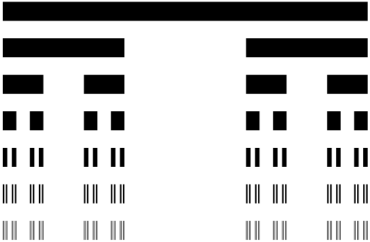
Outline

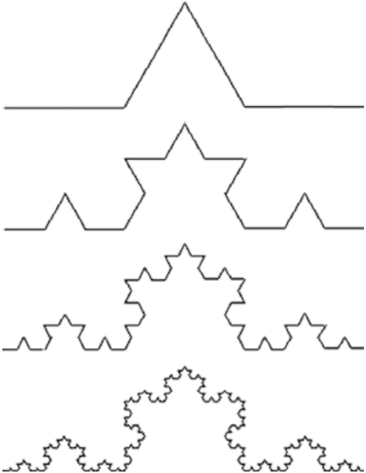
1. Mathematical Background
   - Shannon Entropy
   - Fractals
2. Fractal Image Analysis
   - The Method
   - Use in Astrobiology

Dimension

A finite set of points has dimension zero; a line segment is one-dimensional; a square is two-dimensional; a bowling ball and a slice of pizza and a graduate student are all three-dimensional. This is an intuitive notion of dimension, but not always an adequate one. Some sets are too *weird* to be described well by an integer dimension. For example. . .

Cantor & Koch





Box Dimension

Definition. The *Box Dimension* (or Minkowski-Bouligand dimension) of a set $S \in \mathbb{R}^n$ is the limit (if it exists)

$$D_{\mathrm{box}}(S) = \lim_{\epsilon \to 0} \frac{-\log N(\epsilon)}{\log \epsilon},$$

where $N(\epsilon)$ is the number of square boxes of side length ϵ required to cover the set.

For any finite sequence $\epsilon_0 > \cdots > \epsilon_n$, the best-fit slope through the points $(-\log \epsilon_k, \log N(\epsilon_k))$ is an approximation to D_{box}.

How to Compute $D_{\rm box}$

Notice in particular that, for any $r > 1$, we can set $\epsilon_k = r^{-k}$ and get

$$D_{\rm box}(S) = \lim_{k\to\infty} \frac{-\log N(\epsilon_k)}{\log \epsilon_k} = \lim_{k\to\infty} \frac{\log_r N(r^{-k})}{k}.$$

- For the Cantor Set $\mathcal{C}$, with $r = 3$ we get

$$D_{\rm box}(\mathcal{C}) = \log 2/\log 3.$$

- For the Koch Curve $\mathcal{K}$, it's easier to see if we use circles (of diameter ϵ) for our covering set; then with $r = 3$ we get

$$D_{\rm box}(\mathcal{K}) = \log 4/\log 3.$$

Entropy of a Fractal

This limit, $\lim_{k\to\infty} \frac{\log_r N(r^{-k})}{k}$, has the same form as the per-symbol entropy of an information source! Here, a "message of length k" is just a choice of one of the $N(r^{-k})$ covering boxes of side length r^{-k}.

Accordingly, we define the *entropy of a fractal* to be

$$H(S) = D_{\mathrm{box}}(S) \text{ bits.}$$

This is the (limiting) amount of information produced each time we "zoom in" by a factor of two.

Entropy of Cantor, Koch

- Cantor: $H(\mathcal{C}) = \frac{\log 2}{\log 3}$ bits (per zoom), or since this works for any r,

$$H(\mathcal{C}) = \frac{\log 2}{\log 3} \frac{\text{trits}}{3x\text{-zoom}} = 1 \frac{\text{bit}}{3x\text{-zoom}}.$$

  Indeed, to zoom in on the Cantor Set by a factor of 3, we must choose "left" or "right" — 1 bit of entropy!

- Koch: Similarly,

$$H(\mathcal{K}) = \frac{\log 4}{\log 3} \text{ bits} = 2 \frac{\text{bits}}{3x\text{-zoom}}.$$

Outline

1 Mathematical Background
- Shannon Entropy
- Fractals

2 Fractal Image Analysis
- The Method
- Use in Astrobiology

A Photograph

Consider a black & white photograph. There a couple of natural ways to represent this mathematically:

- As a function, $I : R \subset \mathbb{R}^2 \to \mathbb{R}$.
- For a digital photo, as an $m \times n$ matrix J.

To analyse it as a fractal, we need a set of points in $\mathbb{R}^2$. In fact, each *threshold value t* induces such a set: define

$$I_t^{-1} = \{(x, y) : |I(x, y)| \leq t\},$$

or in the matrix case,

$$J_t^{-1} = \left\{ \left(\frac{i}{m}, \frac{j}{n} \right) : |J_{ij}| \leq t \right\}.$$

Fractal Spectrum

Definition. The *fractal spectrum* of an photograph I over a set of threshold values, $S \subset \mathbb{R}$, is the function $\mathrm{FS} : S \to \mathbb{R}$ given by

$$\mathrm{FS} = D_{\mathrm{box}}(I_t^{-1}).$$

For an image matrix we take an approximation to D_{box}.
Finally, if $S = \{t_1, \ldots, t_m\}$ is finite, we may write

$$\mathrm{FS} = \left(D_{\mathrm{box}}(I_{t_1}^{-1}), \ldots, D_{\mathrm{box}}(I_{t_m}^{-1})\right) \in \mathbb{R}^m,$$

where typically $t_1 < \cdots < t_m$.

Example

Let $J = \begin{bmatrix} 4 & 2 & 1 & 3 \\ 1 & 6 & 1 & 7 \\ 3 & 2 & 8 & 7 \\ 4 & 5 & 5 & 6 \end{bmatrix}$. So with $t = 4$, $J_t^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

To compute $D_{\mathrm{box}}(J_4^{-1})$, write $N(1) = 1$, $N(\frac{1}{2}) = 3$, and $N(\frac{1}{4}) = 8$. The log-log slope comes to approximately 1.585.

Similarly, $D_{\mathrm{box}}(J_2^{-1}) \approx 1.161$ and $D_{\mathrm{box}}(J_6^{-1}) \approx 1.850$

Fractal Excess

A high fractal dimension does not necessarily mean an interesting photograph!
In the matrix case, we can compare this photograph to others of the same size and pixel intensity distribution.
Definition. Let J be an image matrix and let $\mathcal{J}$ be the set of all copies of J with permuted entries. Then the *fractal excess* of J over a threshold set S is the function

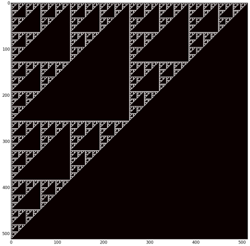
$$\mathrm{FE} = \frac{1}{|\mathcal{J}|}\left(\sum_{J' \in \mathcal{J}} \mathrm{FS}_{J'}\right) - \mathrm{FS}_{J}\,.$$

Again, for finite S, we may write FE as a vector in $\mathbb{R}^m$.

Examples

With J as in the previous example, it turns out that $\mathrm{FE}_J \approx (-10^{-13}, 10^{-13}, -10^{-14})$ over $S = \{2, 4, 6\}$. This is small because there's not much room for anything interesting to happen in a 16 pixel photograph.

In the photograph at right, the dimension is again about 1.585, but computing the fractal excess using the definition would be impractical; instead the image was compared to a single random permutation of itself. The FE is about 0.1.



N.B.

As that last example shows, the numeric values obtained for FE are highly dependent on the resolution of your image.

Outline

1 Mathematical Background
- Shannon Entropy
- Fractals

2 Fractal Image Analysis
- The Method
- Use in Astrobiology

I return now to the paper...

So it looks like they've shown ($p < 0.05$) precisely the opposite of what they set out to show: life exhibits *higher* entropy than visually similar abiotic phenomena!

But this really is due to a confusion of definitions. The fractal entropy the authors measure is not the same as the thermodynamical entropy they discuss in their introduction.

What the authors have shown is some connection between the two; an investigation into that connection would be a logical next step.

Intuitively, it seems they're measuring complexity in some nebulously defined sense: a poorly-shuffled (low entropy) deck of cards still has higher entropy than a fair coin toss, simply because there's more going on.

What Next?

This project was a blast, and raised a lot of new questions that I'd like to pursue at some point.

- How good are these approximations to D_{box} for image matrices?
- How does this new quantity "fractal excess" behave? It clearly depends on the image resolution, but in what way, and can we "normalize" for that somehow? Can we compute the average fractal dimension of a scrambled image combinatorially?
- My reading has hinted at a deep connection between all the various animals called "entropy": information theoretic, classical thermodynamic, and statistical thermodynamic. I'm very curious to learn more.
- What is the connection between the fractal complexity of a physical object and its physical entropy?

Thanks, Edoh!

To conclude, I'd like to thank my advisor Edoh Amiran for his invaluable help and support throughout this project.