From Fractal Geometry to Statistical Fractal

[1]Roberto N. Padua and [2]Mark S. Borres

Abstract

The development from fractal geometry to fractal statistics was established in this paper. Interesting features such as self similarity, scale invariance, and the space-filling property of objects (fractal dimension) of fractal geometry provided an enormous groundwork to build a link towards the statistical paradigm since most data sets are endowed with its non-normal and irregular characteristic. A new probability distribution called fractal distribution was modeled to accommodate this non-normal conforming characteristic of most data sets and sample investigations were presented at the end of the paper.

Keywords: fractal geometry, fractal statistics, fractal dimension, scale- invariance, self- similarity, probability distribution

1.0 Introduction

Fractal is a general term used to describe both the geometry and the processes which exhibit self-similarity, scale invariance, and fractional dimension (Mandelbrot, 1982). Deterministic fractals had been extensively studied and consist essentially of establishing an initiator structure and a generator (Mandelbrot, 1967), Palma (1992), Orbach (1992). The initial structure is transformed by the generator repeatedly at various scales. Common examples of fractals geometric structures include the fractal disks or Cantor sets (dimension d=0.63), the Von Koch curve (dimension d=1.26), Julia sets and others. The applications of the fractal geometry have been enormous; in astronomy, biology and chemistry, data compression, economy, art and music, and weather to name a few. As knowledge seeking progresses, fractals not only gained its grandeur in geometrical sphere, but also in the area of statistics. The study of Padua, et al. convincingly provided the idea on the **strong connection** between fractals and probability theory called statistical fractal (fractal statistics). This is because most real- life situations involve a non-existent mean like dealing with stock market crashes, earthquakes, test scores of students, and others (Selvam (2008)). The non-existence of the mean guarantees a non-existence of the standard deviation and variance hence performing an inferential test (which commonly assumes normality) is impossible. They argued that most of the phenomena that had been modeled using the normal distribution can be more accurately analyzed using statistical fractals instead since most of the data sets exhibit self-similar stochastic patterns (Padua, et.al, 2013).

There still have been few formal studies exemplifying the evolution from fractal geometry to statistical fractal. This study continues to uncover that link from fractal geometry to statistical fractal by showing that the properties of fractal objects (self-similar, scale invariant, heterogeneous, and has fractional dimension) can be linked to the properties exhibited by a random variable in the statistical fractal.

---

[1]Corresponding Fractal Statistics Expert and Consultant
[2]Research Staff, University of San Jose Recoletos

2.0 Fractal Geometry: Self-Similarity, Scale Invariance, Fractional Dimension

Two of the core properties of fractals are self similarity and scale invariance. Formally, self-similarity is defined as a property where a subset, when magnified to the size of the whole, is indistinguishable from the whole (Mandelbrot, 1967). The structure of the shape is made of smaller or bigger copies of itself: same shape but different size. For example, if the Barnsley fern leaf (Figure 1) is closely examined, it is noticed that the every little leaf, part of the bigger one, has a similar shape with the whole fern leaf. So we say that the fern leaf is exhibiting **self- similarity**. This will then suggest that the pattern at the smaller scale is still the same pattern in the bigger scale (magnified) hence **scale invariant.**

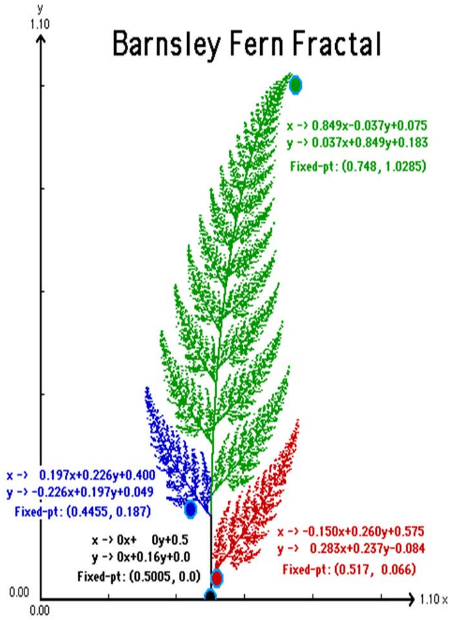


Figure 1

Classical geometry deals with shapes or objects described in integral dimension. A point is 0-dimensional, a line having 1-dimension, a plane figure with 2-dimensions and the 3-dimensional solid. An example below will show how this dimension value is determined.

From a regular square, divide the whole region such that the horizontal and the vertical axes will be partitioned twice, and call the number of partition as r (see Figure 2).

Then, we count the number of replicates (the number of squares) that resulted from the partitioning and call it m.

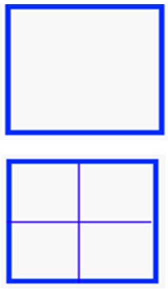


Figure 2

Here, the $r = 2$, and $m = 4$. We have produced 4 similar images (square) by partitioning it (the horizontal and vertical axes respectively) in 2 parts. Now, we give a relationship between the number of images created by the number of partitions and get

$$4 = 2^d,$$

where d is the dimension. To obtain d, we do

$$log\ 4 = dlog\ 2$$

$$d = \frac{log\ 4}{\log 2}$$

$$d = 2$$

or we say 2-dimensional, as expected.

Generally, the dimension of an object having *m* copies of itself and scaled by a factor of *r* is:

$$d = \frac{log\ m}{\log r}$$

However, many phenomena are appropriately described in terms of a dimension between any two dimensions. A straight line has dimension d = 1 and a zigzag of this will have a dimension between one and two for the curve is more than a line but less than a 2-dimensional figure. Here, the dimension manifested is referred to as a **fractional dimension**- a dimension whose value lies between integral values.

We can illustrate this by using one of nature's objects known as the *"snowflake"*.

We construct the *"snowflake"* by drawing a straight line of length 1 unit, call this the **initiator**. Divide this length into 3 equal parts and we will replace the middle segment and replace it with an upside down "V" shape, and now the whole pattern is made up of four line segments This rule is what we termed as the **generator** (as shown in Figure 3).

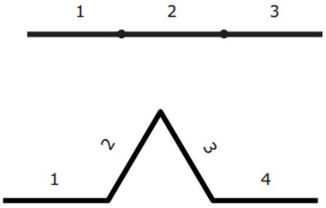
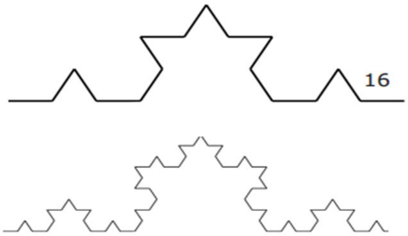


Figure 3

Next, we do the same thing again. Each of those four lines is divided in thirds, and the middle segment is replaced with a "V".

We will then replace each of the line segments with the same pattern again. Eventually, the pattern starts to look like a fractal in nature, such as a coastline, or part of a snowflake (refer to Figure 4).



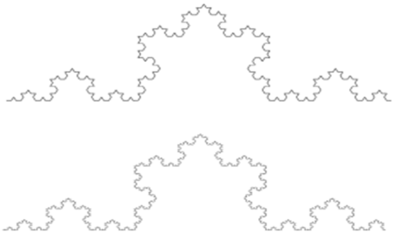


Figure 4. Part of the Snowflake

The fractional dimension of the snowflake is computed to be

$$d = \frac{log\ 4}{\log 3}$$

$$d = 1.26$$

In the same aspect, the dimension can be viewed as the space-filling property of an object in nature. So, the larger the fractal dimension is, the bigger the area that it fills in space, say a cube (d =3) filling more space than a line (d =1).

This nature of the geometrical fractal to accommodate and appropriately treat irregular and randomly structured objects makes simulation to a new branch of statistics probable and we will call this as **Statistical Fractal**.

I. Statistical Fractal: Self-Similarity, Scale Invariance, Fractal Dimension

"The most useful fractals involve chance ...both their regularities and their irregularities are statistical." - Benoit B. Mandelbrot.

While fractal geometry is inherently deterministic since the shape and configuration of the figure are essentially determined by an initiator (figure) and a rule (generator), the resulting object appears to have features that behave like random variables. Naturally, objects like clouds, coastline, landscapes, and even the Fibonacci patterns exhibited by the arrangement of the leaves around the stem of a plant are found to be irregular and non-conforming to normal patterns.

Likewise, data from statistical distributions follow the same pattern of irregularity.

A. Income Distribution:

- There are more people having **low incomes** than **High Incomes.**
- Among those with **low income**, there are more having even lower incomes than higher incomes; same is true for the high income groups.

B. Test Scores

- There are more low test scores than high test scores.
- Among those with low test scores, there are more having even lower test scores

C. Conflicts

- Conflicts arise at various scales: family level, community level, municipal level, Provincial level, regional level, national level, international level;
- The conflict varieties are repeated at each level

D. In Biology

- Organisms trace a path that is known to be fractal and generally obeys the fractal dimension of its ecosystem. This observation is very important in landscape ecology.
- Likewise, the distribution patterns of biological organisms across various trophic levels are fractal e.g. more at the primary producer level than at the top consumer level. The life cycles of organisms (insects to mammals) are also observed to be fractal.

E. In Economics

- Stock market price fluctuations are representative of fractals. Low price fluctuations dominate high price fluctuations.

F. In Catastrophic Events

- Intensity of Earthquakes and the interevent (time it takes for one shake after the other) is found to be non normal and follows a fractal order

G. In Sports

- Statistical data show that the points per game of individual players correspond to an irregular pattern as depicted in most of the case.

For this reason, it will be expedient to develop an axiomatic system that would place fractal geometry in a statistical context. To this end, we begin with a definition:

Definition 1: Let $f: R \rightarrow R$, then f is a **scale invariant** iff

$f(\alpha x) = \alpha^k f(x)$ for some *k*, $\forall\alpha$

Definition 1 suggests that the scale-invariant function f: R->R are the *power functions* which are often used to model events where there are more small values than the large values of X and leads to the following theorem.

Theorem 1: If f(x) is scale-invariant of order k, then

$$f(x) = Ax^k$$

Proof: $f(x) = f(x.1) = x^k.f(1)$

$$= x^k A$$

$$= Ax^k \text{ , where } A = f(1).$$

According to Padua (2013), the particular power- law distribution which will be useful to define invariant probability distribution is given by

$$f(x) = \frac{\lambda-1}{\theta}\left(\frac{x}{\theta}\right)^{-\lambda} \text{ , } where\ x \text{ is a score,}$$

$$\theta < x < \alpha$$

$$\theta = \min\{x_1, x_2, \dots, x_n\}, \qquad 1 < n < \alpha$$

where the exponent of the power distribution, λ, is the fractal dimension of the random variable

X and the corresponding cumulative distribution of X is shown to be

$$F(x) = 1 - \left(\frac{x}{\theta}\right)^{1-\lambda} \qquad \theta < x < \alpha$$

We learned in fractal geometry that the fractional dimension λ represents the space-filling property of an object. Statistically speaking, λ contains information on how irregular and rugged (non-normal) the nature of the probability distribution is where the scores are coming from.

Understanding fractal statistics, which exhibits non-conformity from a normal distribution, is vital since most of the statistical treatments given to these data sets are based on the assumption of normality. Since most of the data sets are fractal by nature, the best model that will analyze the ruggedness of their probability distribution is the backbone of FRACTAL DISTRIBUTION.

The Fractal Dimension λ

In the geometric view, fractal *d*imension is calculated generally as $d = \frac{\log m}{\log r}$ where this amount describes the space filling property of the geometric object. In statistics, the fractal dimension λ can be estimated through the getting the average of the values from the formula:

$$\lambda = 1 - \left(\frac{\log(1-\alpha)}{\log(\frac{x}{\theta})}\right)$$

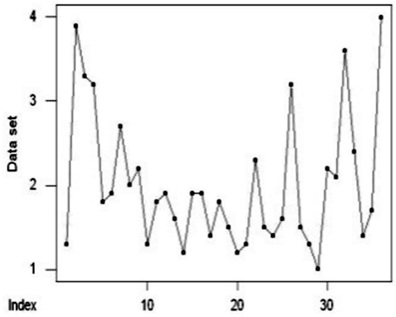
$$\theta = \min\{x_1, x_2, \dots, x_n\}, \qquad 1 < n < \alpha$$

$$\alpha = rank(x)/n$$

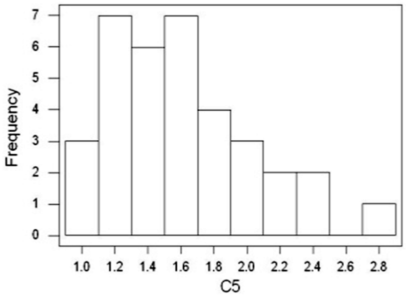
where λ is the amount of information on how irregular the scores are at different scales ($\log(\frac{x}{\theta})$). We note that the lower the scale, the higher the scores are.

A view on λ

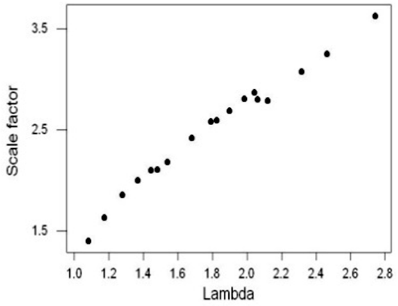
Consider a data set on students' grade in a certain class (Discrete Mathematics-found in appendix A-1). The graphs below indicate the a. time plot series of the data set showing irregularity of spikes b. the histogram of the values of lambda c. the lambda vs. scale graph showing where average lambda is located



a.



b.



c.

Using MiniTab to transform and analyze the data, we come up with

Variable	N	N*	Mean	Median	TrMean	StDev
Lambda	35	1	1.6238	1.5411	1.5969	0.4223

where we calculated λ to be approximately 1.62.

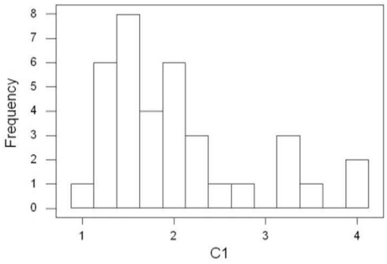
A Glimpse of NON-NORMALITY of Most Data

1. Data Structure Test Scores

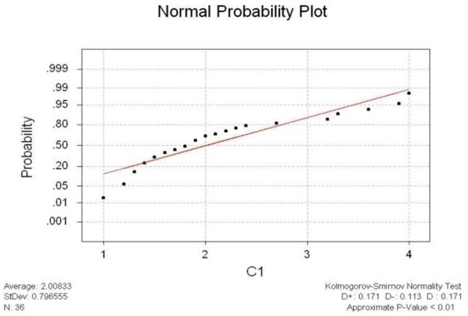
Below are graphs (a. Histogram b. Kolmogorov-Smirnov Test) of data from prelim, midterm, and final grades of students in a Data Structure class of the researcher. 36 students were enrolled in the subject and their grade coverage is for the 2nd semester of SY 2012-2013. All data sets were found to be non-normal.

PRELIM Grades:

A. Histogram

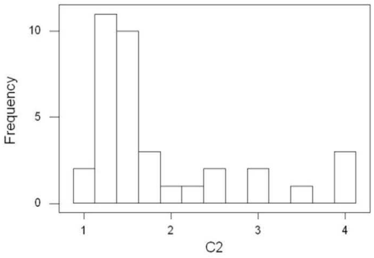


B. Normality Test

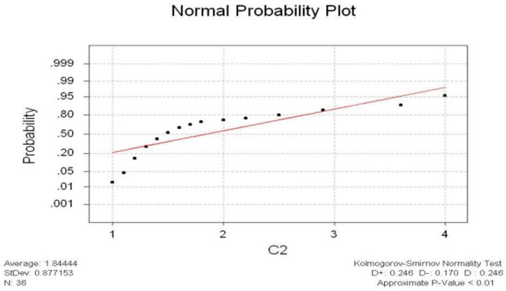


MIDTERM Grades:

A. Histogram

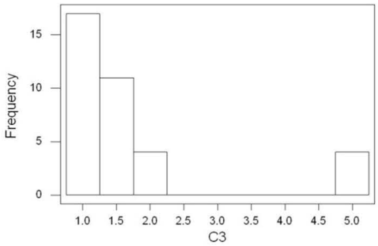


B. Normality Test

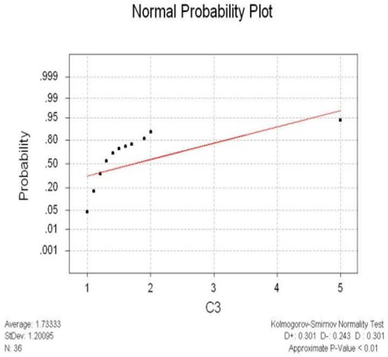


FINAL Grades:

A. Histogram



B. Normality Test



Discussion

From the histograms (a), the raw scores (grades) of students were found to be non-normally distributed. In fact, most of scores lean greater than the mean. The Kolmogorov-Smirnov Tests also indicated p-values which are so small to reject the non-normality of the distributions, and to conclude that the data distributions are really non-normal. Consequently, all analyses that could have been applied to this data set with normal assumptions would be deemed inappropriate.

Fractal Analysis

- Refers to a statistical analysis of fractalobservations based on breaking down the data into fractional or smaller scales (s)
- Given a data set, we first look at the entire ruggedness index (λ). We then assess the ruggedness of the data at each scale (s) by looking at λs.

COMPARISON OF ANALYSES: Data Structure Class Grade Sample

<table>
<tr><th>NORMAL DISTRIBUTION (Assumption)

Classical Analysis (Average: μ)</th><th>NON-NORMAL DISTRIBUTION:

Fractal Analysis(Fractal Dimension: λ)</th></tr>
<tr><td>SUMMARY
<table>
<tr><th>Groups</th><th>Count</th><th>Sum</th><th>Average</th><th>Variance</th></tr>
<tr><td>Prelim</td><td>36</td><td>72.3</td><td>2.008333</td><td>0.6345</td></tr>
<tr><td>Midterm</td><td>36</td><td>66.4</td><td>1.844444</td><td>0.769397</td></tr>
<tr><td>Final</td><td>36</td><td>62.4</td><td>1.733333</td><td>1.442286</td></tr>
</table>
From this result, we can infer that during the prelim term, on the average, the class performed a grade of 2.0 with the deviation of 0.6345. The traditional statistical descriptive analysis smooths out the idea that there are more higher grades than the average or the low ones.</td><td>The computed fractal dimension (ruggedness index), the λ of the prelim grades is approximated at 1.62. Looking at the lambda-scale graph, 1.6 lies in the low level of the scale which means that the irregularity occurs in the high grades: the fractality lies on the higher grades and from here we can do some inferences on the occurrence of this irregularity.</td></tr>
</table>

DISCUSSION:

Data Structure is one of the most fundamental courses of students enrolled in the Information Technology and or the Computer Science. Being an instructor of this subject, the researcher came to wonder about the characteristics the students were showing in terms of their grades in the three terms: irregular. The three terms show a considerable number of students who are within the average/higher but there also a number of students who cluster in the lower grades. (as depicted in the histograms). Having this in mind, to describe the totality of the whole class' performance in terms of the mean grade with respect to the whole class is inappropriate. Instead, what is more fitting is to give description or analysis about the grades with respect to their scales (high grades, average grades, and lower grades).

ACKNOWLEDGMENT

The author would like to thank **Dr. Roberto N. Padua** for the mentoring, and to the anonymous reviewer of the article.

References

Akinyemi, Gbenga M. and Abiddin, Norhasni Zainal. (2013). Human Capital Developments an Interdisciplinary Approach for Individual, Organization Advancement and Economic Improvement", Asian Social Science.

Barnsley, M. (1988). Fractals Everywhere. San Diego, CA: Academic Press Inc.

Borres, M. et al. (2013). Some Results on Multifractal Spectral Analysis. Unpublished manuscript.

D. Rendon. (2003). Wavelet based estimation of the fractal dimension in fBm images, First International IEEE EMBS Conference on Neural Engineering 2003 Conference Proceedings CNE-03, ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8511

Li, Q., et al. (2002).Computer vision based system for apple surface defect detection, Computers and Electronics in Agriculture.

Padua, R., et al. (2013). Statistical Fractal Inference. Unpublished manuscript.

Padua, R., et al. (2013). Statistical Analysis of Fractal Observations: Applications in Education and in Poverty Estimation. Unpublished manuscript.

Sun, W., et al. (2006).Fractal analysis of remotely sensed images: A review of methods and Applications.

Similarity Index: 3%
Grammarly: 99%