The Complexities of Measuring Complexity

Michael Thompson
School of Business: Marketing and International Business
University of Western Sydney

Louise Young
School of Business: Marketing and International Business
University of Western Sydney
Department of Entrepreneurship and Relationship Management, University of Southern Denmark

Competitive Paper
To be presented at IMP Conference, 2012, Rome, Italy

Abstract:

This paper considers alternative ways of conceptualising and measuring complexity in networks. There is a need to include analysis of real world data in the design and verification of models of relationships and network evolution. Often available data is quantitative and survey-based cross sectional or time series data. This paper highlights the problems associated with using statistical analysis to analyse these data and proposes different approaches that conduct analysis of the patterns of change in a way that does not contradict the primary assumptions of statistical analysis.

Keywords: complexity, social system models, information systems

INTRODUCTION

There is increasing recognition that business networks are complex and adaptive (e.g. Holbrook 2003, Wilkinson and Young 2002, 2005, Rand and Rust 2011). However research into the nature and properties of business networks is problematic because rather than top-down directed activity they have properties of bottom-up self-organisation where local actions and interactions taking place guide the nature and evolution of the system. Macro structures and order emerge and are reproduced, or not, over time in a continuous process. Interactions between outcomes of interactions produce unpredictable, non linear systems.

The qualitative and quantitative measurement of complex systems presents theoretical and practical challenges (Gell-Mann 1994a). Systematic case histories and simulations are among the methods recommended to meaningfully research complex business systems as these approaches can capture their dynamic, interconnected nature (Bairstow and Young 2012, Wilkinson and Young 2012,). However much of the research undertaken in business marketing including consideration of B2B relationships and collections of relationships remains quantitative and survey based (Denize and Young 2007).

Survey methods have been argued to be complementary, providing additional insight in multi method studies of business networks (Wong et al 2010). While that may be true, such work relies upon findings emerging via regression-based statistical analysis. This paper highlights the perils of unquestioningly using findings emerging via these standard-practice statistics that are all-too-common in business marketing. The focus for this examination is via consideration of the unsuitability of the underlying assumptions of the measurement methods that are commonly used in the analysis of complex social systems.

To move beyond using complexity primarily as a metaphor, measurement of the nature and evolution of complex systems is important. In the social sciences many of the traditional assumptions and methods of measurement are inadequate for understanding a large part of the behaviour of complex social systems (McKelvey and Andriani 2005). Methods of measurement that capture <u>both</u> the nuances of social systems and the nature of complexity are needed. This requires more than an ad hoc use of complex measurement techniques. The concepts of complex measurement need to be incorporated into the general framework used to measure and analyse social systems as this framework directs many of the assumptions used by social scientists to conduct research (Wilkinson and Young 2002, McKelvey and Andriani 2005, Rissanen 2007).

This paper addresses these issues by considering complexity social systems in terms of the way they function by processing and exchanging information, the complex properties of this and the way these properties need to be measured. This is contrasted to traditional statistical measurement and its assumptions. The paper concludes with a review of developing complex measurement approaches for business systems.

COMPLEX INFORMATION

Social systems are dynamic complex systems consisting of numerous individuals, or agents, connected to each other through a space consisting of networked conduits. These conduits can take many forms, including physical, technological and somewhat abstract socially constructed conduits. Social agents then interact with each other by exchanging information through these conduits. Complex social systems are information processing systems which receive input information from their environment, change this information in some way and then output this information back into the external world. Within the system information is also received and processed in this way and used within the system to guide and direct it. Through this processing of information space-time information, structures emerge and these information structures can be highly complex and contain much information. Work on cellular automata and within information theory highlight that information structures have universal and measurable properties (Langton 1990, Mitchell, Hraber and Crutchfield 1993, Watts and Strogatz 1998, Wolfram 2002, Feldman *et al* 2008, Cook 2009, Sutner 2010).

The properties of complexity are such that even simple rules of interaction can produce complex information structures. And these complex structures often form from their rules in a very counter-intuitive manner, i.e. knowing the rules of interaction does not necessarily give an indication of what kind of information structure will emerge. Also, extremely precise measurement may be needed when dealing with complex systems. Even very small changes to the input or structure of the system can lead to dramatically different results. It follows that imprecise measurements can lead to highly misleading results. Any information lost through the limitations of the measurement process might be highly significant for understanding the emergent properties of a system (Laughlin 2005, Crutchfield *et al* 2009).

Developing an understanding of complex systems is difficult due to limitation of available information about the past (Gell-Mann 1994a). There are also various practical limitations of measurement that restrict the information that can be collected. There are limits to the current theoretical concepts, measures and quantities used to describe complex information (Feldman *et al* 2008, Shalizi 2009, Wiesner *et al* 2009). And there are, of course, also limits to the usefulness of the measurement instruments used to collect information generally (Louviere 2006).

Perhaps the most serious problem for the social sciences is that methods developed for research into complex information are generally not used. Instead the measurement and analysis of complex social systems is mostly carried out using traditional analytical methods which are not adequate for this purpose (Andriani and McKelvey 2009). Traditional statistical methods are often based on assumptions which do not hold in complex social systems and the results they produce can therefore be distorted and misleading. What is more, traditional statistical methods often rely on various aggregation techniques that remove many of the important features of complex behaviour.

In social science and more specifically in business, these methods are creating new methods, of enquiry some of which are and arguably flawed. The assumptions and techniques are used drive research agendas rather than the other way around (Denize and Young 2007) and this creates severe limits on the type of problems which are considered (Andriani and McKelvey 2009, Laughlin 2005). As a result, much of any real system remains hidden from researchers (Borsboom *et al* 2003, Crutchfield *et al* 2009).

TRADITIONAL STATISTICAL APPROACHES

Much of the standard analytics widely used in social and business research are ‘traditional’ statistics (Kvam and Vidakovic 2007, Rissanen 2007, Olson and Delen 2008). These methods are based on assumptions about the nature of data which generally are not met for data associated with complex systems (Geweke 1992, McKelvey and Andriani 2005).

Traditional analytical statistical methods evolved over time out of the need to find signals in noisy data, for instance, in distinguishing results from random data generated by measurement, calculating the effects of different treatments in experiments and determining the relationship between two or more random variables (Mendenhall *et al* 1981, Johnson and Wichern 1992, Weisberg 2005). These methods require certain assumptions about the structure of the data and assumptions about the way data has been generated (Mendenhall *et al* 1981, Weisberg 2005, Wu and Zhang 2006, Rissanen 2007).

These assumptions do not just direct the techniques which are used to analyse data but drive assumptions about the structure of data, how the data has been generated and even the nature of the problems which are considered appropriate for analysis (McKelvey and Andriani 2005, Rissanen 2007). Some of the more important assumptions include "measurement" issues of distribution and structure, such as normality, i.e. the way data is generated, randomness and noise and more general issues of the nature and meaning of analysis and interpretation.

MEASUREMENT ISSUES AND COMPLEXITY

Many standard statistical analysis methods make assumptions about the way in which data is distributed as Kvam and Vidakovic (2007 p1) explain: "Traditional statistical methods are based on parametric assumptions: that is, that the data can be assumed to be generated by some well-known family of distributions, such as normal, exponential, Poisson, and so on. Each of these distributions has one or more parameters". These different families of distributions have different shapes and so indicate different likelihoods that particular values will occur. However, these common assumptions about the way data is distributed are not necessarily correct and this is often the situation with data generated from complex processes (McKelvey and Andriani 2005). These assumptions may be used to make calculations easier rather than because they truly represent data. As Kvam and Vidakovic (2007 p3) note, they are used because they "perfectly convenient" rather than because they are appropriate.

For complex (social) systems' data, many basic assumptions of the classical methods of analysis are extremely questionable (Granger and Orr 1972). One of the frequent assumptions made in the traditional quantitative analysis of social systems is that the data is inherently normally distributed (McKelvey and Andriani 2005). The normal, or Gaussian distribution, is a common way that data is distributed, i.e. in data is clustered around a central value in a particular manner (Frank and Smith 2010).

The assumption of normally distributed data is often justified by the central limit theorem. However, various conditions are required to be met before the mechanisms of the central limit theorem will lead to the production of data that is normally distributed (Mendenhall, Scheaffer and Wackerly 1981) and these conditions generally are not met in complex systems of interacting components (Andriani and McKelvey 2009). One of the key differences between the circumstances applicable to the production of normally distributed data and the behaviour of complex systems is the requirement of independence, that is:

> "In a Gaussian distribution events are assumed to be *in*dependent. Independent events generate normal distributions, which sit at the heart of modern statistics. When events are *inter*dependent, normality in distributions is not the norm"
>
> McKelvey and Andriani (2005 p220)

As the mechanisms within complex systems lead to interdependence, normal distribution statistics cannot be considered in general to be appropriate when dealing with complex social systems. As Andriani and McKelvey (2009 p1053) comment: "The adoption of normal distribution statistics carries a heavy burden of assumptions. Reliance on linearity, randomness, and equilibrium influences how theories are built, how legitimacy is conferred, and how research questions are formulated."

For instance, the mechanisms related to interactions in complex systems can produce data that is distributed in ways other than the normal distribution (Bak *et al* 1987, Barabasi 2003, Andriani and McKelvey 2009). One distribution often produced is the power law distribution. These are found throughout the natural and social sciences (Buchanan 2000) such as in economic data (Kendall 1953, Mandelbrot 1963). Power laws distributions are characterised by a high number of extreme values at one end of the distribution and a single long tail. For example academic paper citation sees a very few papers get a lot of citations and the great majority get zero or one citations. And so if distributions are assumed to be normal when they have longer tails and more extreme values, then the likelihood of extreme events may not be truly represented.This can cause practical as well as research problems as "under many circumstances what is important to most managers are the extremes they face, not the averages" Andriani and McKelvey (2009 p1053).

Therefore, "In the case where the experimenter is not sure about the underlying distribution of the data, statistical techniques are needed which can be applied regardless of the true distribution of the data" (Kvam and Vidakovic 2007 p2). Statistical researchers have developed methods such as nonparametric statistics which do not make such restrictive assumptions about the distribution of data and are therefore robust to a wider variety of data situations (Kvam and Vidakovic 2007). However, although many nonparametric methods offer advantages for a range of real world statistical problems (Kvam and Vidakovic 2007) they can still rely on many of the underlying assumptions of traditional statistics such as independence (Mendenhall *et al* 1981) and are not therefore a fundamentally different approach (Rissanen 2007).

The data emerging from complex processes also has important implications for the fundamental concepts of signal and noise (Rissanen 2007, Andriani and McKelvey 2009). Data generated by complex systems such as complex social systems does not in general meet these assumptions (Geweke 1992, Rissanen 2007). Separating signal and noise in data analysis is inherently difficult (Gell-Mann 1994a) with variability in response outcomes and random components likely to be associated with many factors (Louviere *et al* 2002).

This can lead to fundamental problems for traditional statistical modelling when the underlying assumptions of the methods are not met as there is no clear way of partitioning the information into either signal or noise (Rissanen 2007): Traditional statistical modelling methods assumes clearly distinguishable signals (independent variables that are usually assumed deterministic and predictable) and noise (chaotic elements) but in complex social systems these are likely to be blurred (Andriani and McKelvey 2009).

These problems show the importance of considering how data is produced and "It is often useful to think about the data on hand as the product of a specific data-generation process, also sometimes called a data-generation mechanism." (Berk 2008 p9). However, exactly what mechanisms lead to what distribution types is still not fully understood (Frank and Smith 2010) and "In general… we do not have enough knowledge of the machinery that generates the data to convert it into a probability distribution" (Berk 2008 p9). This leads to problems with the fundamental conceptual basis used in traditional statistics to analyse and explain the origins of data as they are reliant on assumptions of data distribution and structure.

ANALYSIS, INTERPRETATION AND COMPLEXITY

Analysis of data and interpretation of statistical findings e.g. statistical confidence and the broad choices made about which statistical approaches to use also differ when data is from complex systems. In traditional statistics a great deal of analysis makes use of aggregation measures such as means and variances to summarize data (Frank and Smith 2010). These types of aggregates are often incorporated in the calculations of many statistical methods such as various regression and multivariate techniques (Mendenhall *et al* 1981, Johnson and Wichern 1992, Weisberg 2005). As already indicated if data is not normally distributed then means and variances cannot be interpreted to have the same behaviour that is ascribed to them under the assumption of normality (McKelvey and Andriani 2005). Means and variances become unstable and unsuitable as the assumptions of independence are not met in complex social systems of interdependent agents (Andriani and McKelvey 2009). Aggregation methods can "miss some key drivers of social systems" (Miller and Page 2004 p9) including things such as time dynamics and agent level interactions.

As basic measures, such as means and variances, are parts of other methods, problems with them leads to substantial problems with interpretation of the methods they are part of. For example, the statistical significance of a result and the width of confidence intervals around a result are utilized in many statistical techniques including regression and multivariate methods (Mendenhall *et al* 1981, Johnson and Wichern 1992, Weisberg 2005). And, if the validity of these confidence measures is in doubt, it casts doubt on the validity of the results and the theories and methods that have used them (Kvam and Vidakovic 2007, Andriani and McKelvey 2009). In practice, many researchers rely on these normal based methods for confidence intervals and measures of statistical significance as well as using multivariate methods that are built upon them (McKelvey and Andriani 2005, Kvam and Vidakovic 2007).

The inherent uncertainty in some data structures, such as those created by complex systems, also creates problems with the fundamental concept of hypothesis testing as used in traditional statistics. Not only can the results of any hypothesis test be misleading, but there are problems concerning what concepts are actually being tested, as Berk (2008) comments:

> "Even under ideal circumstances, however, statistical inference can be problematic. If the goal is to construct confidence intervals, very little may be known about the relevant sampling distribution. If a priori ignorance prevails about the f(X)) [the model of the data relationships] as well, there may be no sensible hypotheses to test that were posed before the data were examined. Hypotheses generated as part of the data analysis process violate a key assumption of statistical tests. The computed p-values [measure of statistical uncertainty] will likely be too small." Berk (2008 p330)

Traditional statistics not only make assumptions about the nature of data but also make assumptions about the type of results that are expected, that is, “Traditional statistical analysis involves an approach that is usually directed, in that a specific set of expected outcomes exists. This approach is referred to as *supervised* (hypothesis development and testing).” (Olson and Delen 2008 p4). However, this is not the only approach to statistical analysis and other approaches seek to explore the data using less assumptions in a “spirit of knowledge discovery” (Olson and Delen 2008 p4). These exploratory methods are sometimes much better suited to the analysis of data generated by complex processes (Berk 2008) as complex data can be generated by many different mechanisms producing many different types of data (Frank and Smith 2010).

Another area of concern is that researchers change data to meet the assumptions of the framework and techniques of traditional statistics. For instance, many quantitative organizational researchers “are trained via Gaussian statistics to go to great lengths to configure their data to fit the requirements of linear regression models and related statistical methods” (Andriani and McKelvey 2009 p1053). This changing of data is sometimes justified by “the so-called “robustness” enhancement techniques described in standard econometric textbooks” (Andriani and McKelvey 2009 p1068). These methods at times “delete outliers on the assumption that they are all errors and anomalies” (Andriani and McKelvey 2009 p1067), whereas these data points may contain genuine and important information (Andriani and McKelvey 2009). This is a particular problem with complex systems which can produce extreme values that should be considered part of the genuine information. Unfortunately, by taking this approach and “adapting the data to the tools rather than vice versa, we limit the scientific questions that we ask to those answerable by the tools we use” (Weiss 2005 p113). In dealing with complex systems this has the potential to remove much of the information which may be important for both researchers and practitioners (McKelvey and Andriani 2005).

A common approach researchers use to gain some control over distributions and data structures is to construct statistically designed experiments (Montgomery 1991). Methods based around experimental design concepts such as conjoint analysis are commonly used in market research (Carson *et al* 1994, Louviere *et al* 2010). However, in order to gain this control over the data usually requires removing the interactions between agents and the complex processes which occur in real world social systems. So, an experimental design approach helps to produce results with increased precision and more reliable statistical confidence but results in reduced meaning for representing situations that occur in real world social systems.

The traditional statistical assumptions also have a deeper impact on the way systems are perceived and studied as, “Additional assumptions ... often implicit, concern evolutionary gradualism and equilibrium” (Andriani and McKelvey 2009 p1065). The methods and assumptions of traditional statistics are often based around systems behaving in smooth and predicable ways which can lead to a ‘linear’ view that affects what is measured, how the interaction between variables is conceptualized, and which methods are used for analysis (Arthur 1995).

However, assumptions of smoothness and equilibrium cannot be made for complex systems. What is more, it is sometimes this dynamic behaviour that is of most interest in social systems (Miller and Page 2004). Measuring these dynamic patterns is therefore very important, however, this is not currently done well in the analysis of complex systems (Crutchfield 2009) and often is not done at all in traditional statistics. Methods that are heavily reliant on assumptions about the patterns in data, such as traditional statistics, are not suited to finding patterns in data as Crutchfield (1994a) comments:

> "detected patterns are often *assumed* implicitly by analysts via the statistics they select to confirm the patterns' existence in experimental data. The obvious consequence is that "structure" goes unseen due to an observer's biases…. It is rarely, if ever, the case that the appropriate notion of pattern is extracted from the phenomenon itself using minimally-biased discovery procedures. Briefly stated, in the realm of pattern formation "patterns" are guessed and then verified."
>
> Crutchfield (1994a p3)

What is more, the dynamic patterns in complex systems can themselves change, such as in phase changes, and can also be highly unstable around phase change boundary regions, making the measurement of patterns in real data even more difficult and even further outside of the capabilities of traditional statistics. And of course, concepts such as statistical confidence intervals lose any real meaning in unstable environments.

COMPLEX INFORMATION MEASUREMENT

This discussion has highlighted that complex phenomena is poorly represented by scales constructed using statistical assumptions and the analysis of those scales. Researchers throughout the different areas of complexity science are developing a general theoretical framework and a variety of alternative methods for analysing complex information (Gell-Mann 1994a, Crutchfield 1994a). A variety of methods are needed. For instance, Crutchfield *et al* (2009) consider:

> "The tools are (i) dynamical systems theory with its emphasis on the role of time and on the geometric structures underlying the increase in complexity during a system's time evolution, (ii) the notions of mechanism and structure inherent in computation theory, and (iii) inductive inference as a statistical framework in which to detect and innovate new representations."
>
> Crutchfield *et al* (2009 p10)

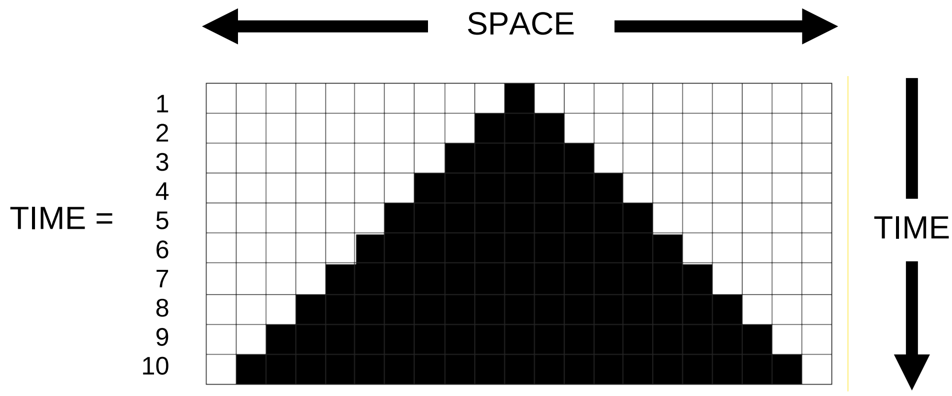
Understanding what the nature of complexity is, is difficult (Gell-Mann 1994a). It follows that measuring complexity is also difficult. There is no agreed upon measure(s) of complexity (Wiesner *et al* 2009). This is because there are many details within complex information (Murray Gell-Mann 1995). As a result there are a large number of measures of aspects of complexity including:

> "... effective measure complexity..., statistical complexity..., logical depth..., thermodynamic depth..., effective complexity..., to mention a few. Some measure structure as expected..., some only randomness... . Some are computable... , others are not... . And most often they are completely unrelated quantities."
>
> Wiesner *et al* (2009 p1)

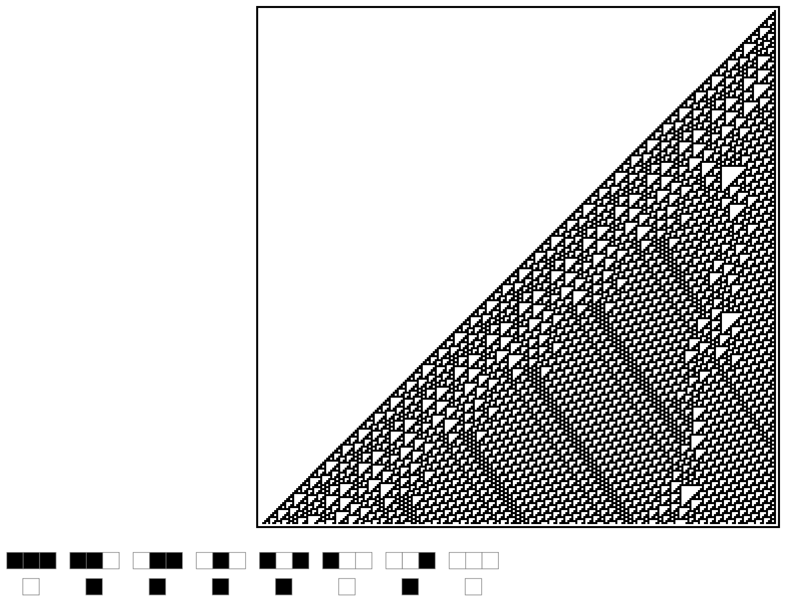
This plethora of different measures has been criticised for its lack of coherence and usefulness and have been described as "just quantification for quantification's own sweet sake" Shalizi (2009 p1).

This discussion indicates that measuring all aspects of complexity is not technically feasible, is unnecessary in many practical applications, and hence is counterproductive. Rather, effective measurement of complexity involves seeking to measure the most important classes of behaviour found in complex information. The literature of cellular automata (Langton 1990, Mitchell, Hraber and Crutchfield 1993, Watts and Strogatz 1998, Wolfram 2002, Feldman *et al* 2008, Cook 2009, Sutner 2010) shows how complex information is comprised of various structural characteristics as well as random behaviour. Understanding complex information, therefore, requires understanding the relationship between these different types of structure, i.e. the ordered part of nature and random behaviour (Crutchfield 1994a).

Figure 1 illustrates a simple version of this, using cellular automata. The row represents all the agents in a system at a point in time. Each column represents an agent over time. Here black cells represent one state and white another. The system of agents is represented over numerous time steps where the interaction between agents causes changes in subsequent states.



<u>Figure 2</u>: Edge of Chaos



In the system depicted in Figure 1 there is a single “rule”, if an agent is in a white state and proximate (immediately adjacent) to an agent who is black state, they switch to black in the next time frame. This forms patterns in space-time as displayed in Figure 1. Other observed structures are different and more obviously contain structure and random elements, as indicated in Figure 2. Here there 256 time periods and a larger number of agents so patterns can be seen more clearly. As with Figure 1 there was a single black cell in period one. There are now more rules for changing from black to white and vice versa. These are at the bottom of the figure, e.g. the first rule states that if an agent is in a black state in a period and has agents also in black states on either side, they will change to a white state in the following period.

These lead to a structure which has both domains of ordered patterns and noise. Random behaviour is generated within complex social systems, including behaviour like traffic jams or certain aspects of human behaviour (Wolfram 2006). It is conceptualized as lying at the other extreme from order, with complexity (edge of chaos) existing somewhere between them (Gell-Mann 1995) and containing elements of both. This is depicted in Figure 2 where patterns can be seen to flip quite abruptly.

One of the advantages of randomness as a measure of complexity is that many different measures attempting to capture quantifications of randomness have been developed (Ebrahimi *et al* 1999a, Ebrahimi et al 1999b). This then enables the quantitative measurement of some notion of complexity to be made. “The most basic statistical test for randomness is the runs test” (Chatterjee *et al* 2000 p658). A run is a sequence of events such as series of coin tosses where there is a series of heads. Runs can be of different lengths. For instance, tossing a coin might result in long runs of heads or many smaller runs of heads and tails. The behaviour of runs gives an indication of the level of randomness so, “A very small or a very large number of runs in a sequence would indicate non-randomness” (Mendenhall *et al* 1981 p597). For example a large number of runs may indicate a cyclical pattern (Freund 1984).

In practice, many measures of randomness-related concepts are actually measures of variation or spread in the data. There are many measures of variation including “standard deviation, mean deviation, median absolute deviation, mean difference, range, interquartile distance and linear functions of order statistics” (David 1998 p368). Variance is the most popular of these (Yitzhaki 2003). However, this does not mean that the variance is the only measure that could be used or that it is always the best measure (Ebrahimi *et al* 1999a). The variance is inherently related to the normal distribution (Mendenhall *et al* 1981). As all normal distributions have the same inherent shape, the spread of any normal distribution can be specified just by the variance. However, if the data is not normally distributed then the meaning of the variance is less clear (Granger and Orr 1972).

One measure of variation that has particular importance for complexity research is entropy (Langton 1990, Gell-Mann 1994a). Entropy is found in many fields of science including statistics and machine learning (Hausser and Strimmer 2009) and its use is growing rapidly (Ebrahimi *et al* 1999a). Entropy and variance are somewhat different. Variance is a measure of spread from the mean of a distribution, whereas entropy measures the spread across the distribution. This means that entropy measures the “diffuseness of the density irrespective of the location(s) of concentration” (Ebrahimi *et al* 1999b p319).

Entropy offers advantages over variance in many situations, for instance, when dealing with data that cannot be assumed to be normally distributed (Ebrahimi *et al* 1999b). A formula for the entropy (Shannon 1948) is given here where the entropy $H(X)$ is related to the probability $p(x_i)$ of the different values of x occurring:

$$H(X) = -\sum_{i}^{n} p(x_i)\log p(x_i)$$

In a curve describing the distribution of data, such as a probability density curve, the height of the curve is related to probabilities of values occurring, and the higher the curve the higher the probability. And so, as the entropy is related to probability of the different values of *x* occurring, the entropy is a measure of the shape of the distribution. The entropy measures whether distribution shapes are evenly spread or concentrated around some values. So the entropy measures how close a distribution is to the flat uniform distribution, with the value of entropy increasing as the distribution approaches the uniform distribution (Ebrahimi *et al* 1999a). As the entropy increases, "the concentration of probabilities decreases" (Ebrahimi *et al* 1999a p3). A more concentrated probability distribution provides more information about where a value is likely to be; a uniform distribution gives no information that one value is any more likely to occur than another (Conrad 2011).

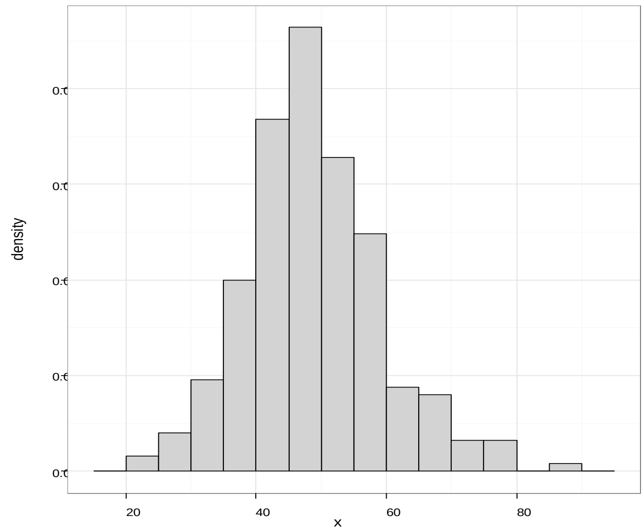
This concept might be explained by considering an intuitive example. Say a visitor to a university is seeking out a particular researcher. If the visitor is told that the researcher is somewhere on campus but no one is sure where, the information is not much use to the visitor. The probability of the researcher's location is spread uniformly across the entire campus and so the entropy is high. However, if the visitor is told that the researcher is probably lurking somewhere around the Marketing School's coffee machine, then the visitor has more information and a greater chance of success. The entropy is lower.

The entropy, therefore, gives a measure that indicates the amount of information, that is, "entropy can be regarded as a measure of ignorance" (Gell-Mann 1994a p219). Entropy-related measures are widely used in information theory (Gell-Mann1994a, Ebrahimi *et al* 1999a) and in the study of complex information (Langton 1990, Gell-Mann 1994a Feldman et al 2008). What is more, as social systems are complex systems and produce and process complex information, entropy can be used to measure and describe features of social systems (Marks 1998).

However, while the variance can be estimated with a simple formula, estimating the entropy is not as straightforward. There are problems with estimating entropy from real data including problems with bias, variation and with small sample sizes (Pincus 1991, Hall and Morton 1993, Hausser and Strimmer 2009). There are a variety of different methods used to estimate entropy and these have advantages and disadvantages, behaving differently with different types of data (Soofi and Retzer 2002, Hausser and Strimmer 2009).

Histograms, such as in the example displayed in Figure 3, are a common method of studying the shape of distributions. As entropy is concerned with the shape of distributions, some methods of estimating entropy are conceptually related to describing a distribution with a histogram (Hall and Morton 1993, Victor 2002). However histogram-related approaches can have problems in the estimation of the shape of the distribution. For instance, the width of the histogram bins changes the shape of distribution as bins that are too wide, lose detail and bins that are too narrow lead to problems with low sample numbers. This, as well as problems with poorly categorizing data, can cause problems with calculating the entropy measure (Victor 2002, Miśkiewicz 2010).

Figure 3: Estimating entropy from histograms



As mentioned earlier, a measure of complexity needs to consider both randomness and order. Historically there has been a focus on randomness to measure complexity, "Eventually, however, it was realized that measures of randomness do not capture the property of organization." (Feldman *et al* 2008, p1). There is also need to capture the properties of "a system's complexity— its organization, structure, memory, regularity, symmetry, pattern, and so on" (Feldman *et al* 2008 p.2).

Some researchers have concluded that a single measure of organisation will not be sufficient and multiple measures are required (Marks 1998). Others seek to develop a summary measure. Various measures have been used by researchers to study structure such as a measure known as the 'excess entropy' which "can serve as a measure of global structure or correlation present in the system" (Feldman *et al* 2008 p4). Various approaches to measuring structure include trajectory and entropy of a series which in turn form the basis of other, more complicated measures.

TRAJECTORY

Measuring structure can be informed by the information stored in the patterns of data that form in space and time (Crutchfield 1994a, Wolfram 2002). Data form paths, or "series" as they are known, through space or time or both (Denis and Crémoux 2002). Time series (Chatfield 1989, Schelter, Winterhalder and Timmer 2007) are commonly considered but a series of data can form through space or some other dimensions as well, for instance, spatial series are studied in geology where patterns in rocks and soil are analysed (Denis and Crémoux 2002). And as demonstrated in the cellular automata examples (Figures 1 and 2), data can form series though both space and time.

Time series are commonly used in the social sciences to measure the change over time of a system. For instance, "the need to assess subtle, potentially exploitable changes in serial structure is paramount in the analysis of financial and econometric data" (Pincus 2008 p329).

And time series are used to help understand the internal processes occurring in social systems, as Dooley and Van de Ven (1999) discuss:

> "Studies of organizational processes can yield observations in the form of event time series that can be analyzed to determine … dynamic patterns. These different patterns each imply different underlying generative mechanisms and hence, different process theories."
>
> (Dooley and Van de Ven 1999 p 358)

However, as with other methods of traditional statistics, traditional time series analysis often relies on assumptions and a general framework that is not appropriate for analysing data from complex systems (Chatfield 1989, Ramos *et al* 2009) including reliance on aggregated data, and the assumption of independence of agents (i.e. there may be an assumption that events are correlated within agents but agents do not affect other agents) (Wu and Zhang 2006).

And as with other traditional statistical methods, traditional time series analysis also often makes assumptions about data behaving smoothly. Time series data is then treated as having predictable trends and seasonal variations while other structures in the data are classified as noise (Chatfield 1989). Therefore, different methods are being developed in order to gather increased information from time series data (Balestrino *et al* 2008, Pincus 2008).

One approach is to analyse the shape of time series (Maragos and Potamianos 1995, Keogh and Pazzani 1998, Liao 2005, Wang, Smith and Hyndman 2006). These methods have the advantage of not being overly reliant on traditional assumptions but classify time series by how they group together in clusters based on shape measures. Similar methods designed to analyse short time series have been developed (Todorovski *et al* 2002, Möller-Levet *et al* 2003, Möller-Levet *et al* 2005) and this helps to deal with the practical problem of limited data points.

A general class of approaches considers that the large scale trends and the small scale fluctuations are all caused by related underlying generative mechanisms and seeks to understand series data from this perspective (Kendall and Bradford Hill 1953, DuPain, Kamae, and Mendés France 1986).

ENTROPY OF A SERIES

A related approach attempts to measure the entropy of a series (Denis and Crémoux 2002, Balestrino *et al* 2007, Pincus 2008, Sabbione and Velis 2010). Earlier, the concept of entropy as a measure of randomness in a data was discussed. The concept of entropy can also be applied to a series of data such as a spatial or time series and has the same basic meaning, as Denis and Crémoux (2002 p899) comment: "Whatever the context, the interpretation of entropy as a measure of an uncertainty or as a quantity representing the state of ignorance of the outcome is straightforward.".

As with the other applications of entropy, the analysis of series data using entropy-related approaches provides additional information about structure when compared with the use of aggregates and traditional time series methods as Pincus (2008) explains:

> “Although analysts typically track shifts in mean levels, variability, and (often) the autocorrelation structure, in many instances a shift in an “ensemble amount of randomness” may provide a critical indicator of asset or market status. Despite this recognition, formulas to directly quantify an aggregate “extent of randomness” have not been routinely utilized in market analyses... Thus in many settings, subtle changes in serial structure would largely remain undetected, unquantified, and/or not acted upon.”
>
> Pincus (2008 p329)

In an intuitive sense the entropy of a series is a measure of its randomness and “Roughly speaking, the entropy of a curve is 0 when the curve is a straight line, and increases as the curve becomes more ‘wiggly’” (Balestrino *et al* 2008 p1157).

As with the calculation of entropy in a distribution, the calculation of entropy of a series is not always straightforward (Miśkiewicz 2010). Different entropy-like measures of time series have been proposed (DuPain, Kamae, and Mendés France 1986, Allouche and Maillard-Teyssier 2010) such as the ‘approximate entropy’ (Pincus 1991, 1995, 2008). One simple method that provides an intuitive understanding is offered by Denis and Crémoux (2002 p904) who consider “there is an evident connection between the classical cumulative sum of absolute values of first differences and the entropy: the entropy of a time or spatial series is the slope of the cumulative sum of the absolute difference”. The measure of the *differences* as used by Denis and Crémoux (2002) is a fundamental concept for understanding the behaviour of a time or other series and a variety of measures are based on differences (von Neumann *et al* 1941, Chatfield 1989).

BASIC SERIES MEASURES

As discussed previously, measures of dispersion such as variance and entropy are based on the difference between data values. These measures of dispersion are often used in situations where there is no sequential ordering of the data and the dispersion is usually measured around the mean or between values. However, when quantities vary over time or some other dimension, an added complication is introduced in that each successive measure is related to the previous measure. The measures are therefore not independent of one another but are related (Weisberg 1985). Using standard methods of calculating variance in such circumstances may lead to misleading results (von Neumann *et al* 1941).

The understanding of series data can be enhanced by making use of the basic concept of a difference. Series data contains additional information in the form of the sequential ordering of data points along time or a spatial dimension. And so one of the simplest methods used to measure how quantities change over time or space is to measure the difference between successive measures (Chatfield 1989).

This has practical application for measurements that move through time or some other dimension. Interestingly, one of the early groups to have used differences to solve their difficulties with the measurement of trajectories through time was concerned with the ballistics of artillery shells:

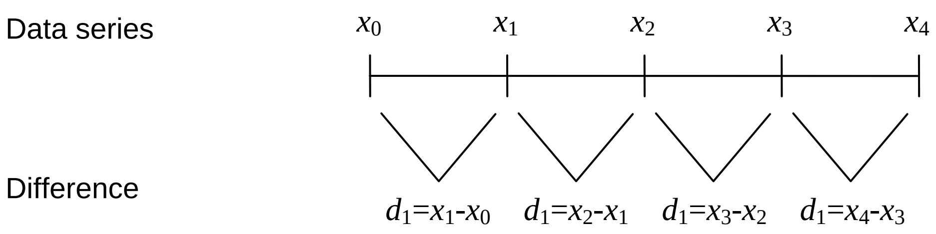
> “The usefulness of the differences between successive observations only appears to have been realized first by ballisticians, who faced with the problem of minimizing

effects due to wind variation, heat and wear in measuring the dispersion of the distance travelled by a shell."

von Neumann *et al* (1941 p154)

Difference values are calculated by simply measuring the difference between each successive measure as demonstrated in Figure 4.

Figure 4: Difference calculation (taking the differences between the nearest values)

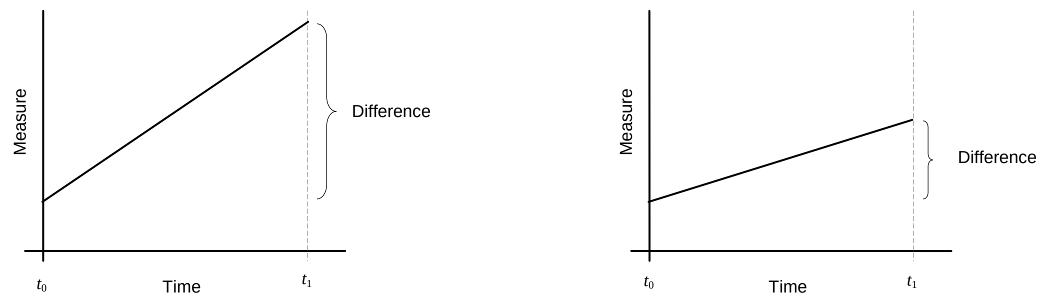


The importance of difference values means they are often expressed as formulas (Milne-Thomson 1933, Weisstein 2010). Here, in an example, the difference between the current value and the previous value for a function f at time t is given by the backward difference operator defined as:

$$\nabla f_t \equiv f_t - f_{t-1}$$

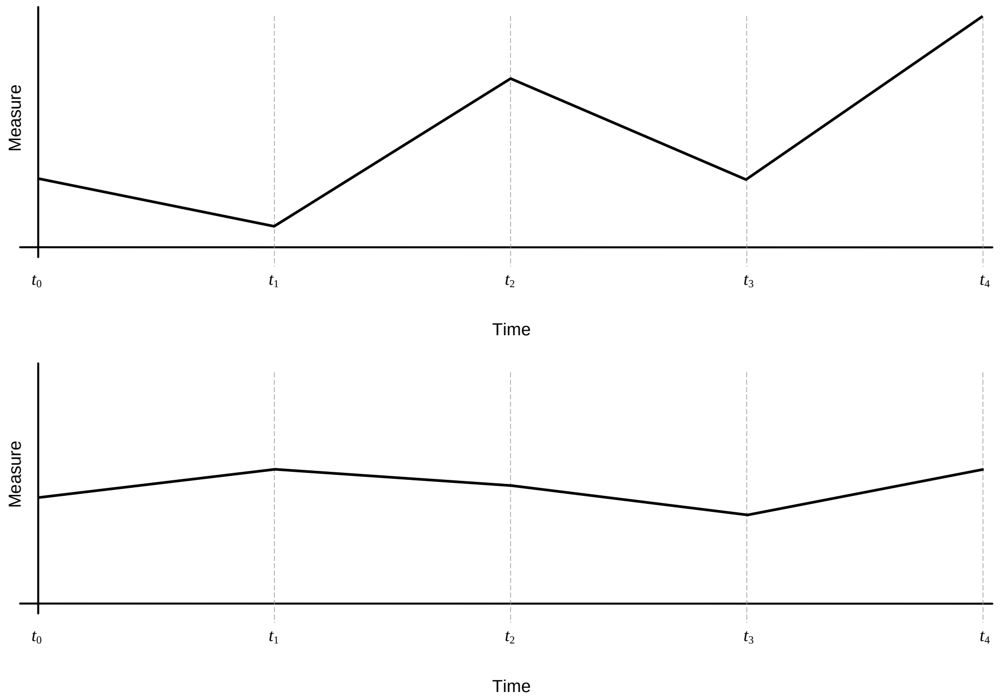
The use of difference has interesting physical interpretations for trajectories as demonstrated in Figure 5. The difference of the measure between t_1 and t_0 is related to the slope and thus represents the rate of change of the measure over time. If the measure was position, the difference would be related to velocity, that is, the rate of change of position. As can be seen below, the graph on the left has a greater slope and hence a greater rate of change than the graph on the right.

Figure 5: Difference Measures over Time



Series with greater successive differences have more change. This is in itself an important measure, however, assuming the change is not mostly in the one direction, then higher differences may also indicates greater variation in the series, and hence more entropy like behaviour. This is demonstrated in Figure 6.

Figure 6: Time fluctuations and entropy-like behaviour.



Different methods can be used to study the behaviour of a series. For instance, histograms of differences can be constructed (Stanley *et al* 2000). Also, summary measures of variation can be calculated, such as the entropy for a series, as described earlier.

One of the earliest and most basic measures of variation of a series is based on the concept of differences and is the ‘mean successive difference’ used by Cranz and Becker (1921 p383) in their work on ballistics, where: “we take the successive differences…without any regard to their signs, and reckon the average value of these differences”:

$$E_d = \frac{\sum_{i=1}^{n-1} |x_{i+1} - x_i|}{n-1}$$

The absolute value of the differences between each successive observation of x is added up and divided by the total number of differences which is n -1. The order through time is described by the subscript i.

There are many other methods for making such measurements (Gasser *et al* 1986, Hall *et al* 1990, Dette *et al* 1998). Von Neumann *et al* (1941) uses a very closely related method of the 'mean square successive difference' which only differs by making use of squared differences instead of absolute differences. However, there are advantages in using absolute values and not squaring, for instance "it is well known that the cumulative sum of absolute differences is less affected by random noise or trend than a sum of square of differences" (Denis and Crémoux 2002 p904).

Differences have intuitive meaning which can be used to study systems in the social sciences. As mentioned, if position is being measured the difference is related to the velocity. In the social sciences, a traditional cross sectional analysis only gives an indication of what value a social agent has at a particular point in time. A difference gives a measure of how that agent is changing its value over time. And, of course, more detailed measures can be constructed out of these basic measurement building blocks.

Higher order differences can be calculated, that is, the difference of the differences. The next difference would correspond to the rate of change in velocity, which is the acceleration. These higher order differences have less intuitive meaning. However, the third order difference is used in many applications. This is the rate of change of acceleration and is known as the "jerk" or "jolt". Gasser *et al* (1986) propose a second order difference measure and Hall *et al* (1990) developed a generalized higher order difference measure.

CONCLUSION: APPLICATIONS OF COMPLEX DATA ANALYSIS

The kinds of complex measures of analysis described above can be used to measure the processes and behaviours of complex systems and have the potential to be extended. Returning to the conceptualization of complex systems as information processes discussed earlier, it is possible to envisage complex methods of analysis that could be used to measure and link the structure of input information, the way systems change information and the structure of output information (Wuensche 1998, Feldman *et al* 2008).

The way information is changed is central. Measures of order and randomness such as entropy are important for understanding the way complex systems process information, for instance, stability is better for storing information and instability for processing of information (Crutchfield 1994a, Lloyd 2000). And so "good" complexity measures will "capture a system's intrinsic computation: how a system stores, organizes, and transforms information" Feldman *et al* (2008 p2).

Thus these complex measures can give insight into the way systems behave and solve problems. For example, as considered earlier, social systems such as businesses and markets can be described by various characteristics such as randomness and clustering (order). The above discussion showed how complex analysis methods can measure these types of structural characteristics in data. Also as discussed, the dynamic behaviour of social systems is an important consideration. For instance, the level of turbulence in a market can be a key characteristic (Wilkinson and Young 2005). And as demonstrated, the complex analysis approach also has the capability to measure and quantify aspects of dynamic space-time behaviour.

However, caution must be exercised in developing and using complex information metrics As with traditional statistics, many of these methods are highly technical and not always useful (Weisberg 2005, Shalizi 2009). Therefore, the methods of complex analysis need to be used within the context of an appropriate framework so that researchers can understand the fundamental principles of both complex data and the methods used to analyse that data. Without such a framework, there is the potential to produce misleading or useless results (Shalizi 2009).

While this paper has demonstrated that a framework of traditional statistics is not appropriate, it is possible that some of the techniques and measures of traditional statistics might be suitable in an appropriate framework to effectively measure the characteristics of complex systems. For instance, although entropy might be a more ideal measure of dispersion, variance still provides a measure of dispersion which may be more practical and still acceptable in some situations.

It is beyond the scope of this paper to present such a framework, however this work highlights that such a framework would appropriately include conceptualization of social systems as complex information structures possibly modelled or visualized as cellular automata. In addition context-appropriate measures of noise, structure and evolution over time and space that are well grounded in the fundamental principles of complex data analysis should be included. From this foundation, more detailed methods of complex analysis can be built.

While researchers in complexity have made progress, there is still much to do. Practical measures of complexity are lacking:

> "Despite the creation of information theory half a century ago with its formal measures of information, entropy, and the Kullback-Leibler distance or the relative entropy, there have been serious difficulties in applying them to make exact the idea of information extraction for model building."
>
> Rissanen (2007 p44)

Future research must address both the construction of measures of complexity and ways of making this useful for researchers in a range of disciplines (Marks 1998). For complex social systems this may prove particularly difficult. There is movement in the right direction. However while pockets of knowledge exist into various aspects of this problem, so far these are not meaningfully connected. This paper takes a step towards linking these and future work will present the framework and ways of using it to analyse complex data.

Events such as this IMP special session further highlight nature of complexity in business and offer solutions such as the myriad of roles that simulations can play. The BNAS project with its goals of developing a user-friendly platform for developing models of complex social systems such that those who do not have the requisite programming skills can participate in this kind of research is another possible way forward (see Wilkinson et al 2011, Wilkinson and Young 2011). Work considering the dynamics of social networks (e.g. the Melnet group at the University of Melbourne) and the study of the quantification of co-occurrence of words in spoken and written discourse and its evolution (e.g. Leximancer, see Smith and Humphreys 2006) are further examples of ways the researchers are seeking to develop the context-specific metrics that deal with complexity.

BIBLIOGRAPHY

Allouche, Jean-Paul and Maillard-Teyssier, Laurence (2010) “Inconstancy of finite and infinite sequences” Preprint submitted to Elsevier, *arXiv:0910.1173v2* [math.DG]

Andriani, Pierpaolo and McKelvey, Bill (2009) “From Gaussian to Paretian Thinking: Causes and Implications of Power Laws in Organizations” *OrganizationScience,* Vol. 20, No. 6, November–December, pp. 1053–1071

Arthur, W. Brian (1995) “Complexity in Economic and Financial Markets”, *Complexity*, 1, 20-25

Bairstow, N and Young, L (2012) How Channels Evolve, a Historical Explanation, *Industrial Marketing Management*.

Bak, Per, Tang, Chao and Wiesenfeld, Kurt (1987) “Self-Organized Criticality: An Explanation of the 1/f noise” *Physical Review Letters,* 59, 381–384

Balestrino, Aldo; Caiti, Andrea and Crisostomi, Emanuele (2007) “A Classification of Nonlinear Systems: An Entropy Based Approach”, *Chemical Engineering Transactions*, vol. 11, pp. 119-124

Balestrino, Aldo; Caiti, Andrea; Crisostomi, Emanuele and Grioli, Giorgio (2008) “A Generalised Entropy of Curves: An Approach to the Analysis of Dynamical Systems” *Proceedings of the 47th IEEE Conference on Decision and Control* Cancun, Mexico, Dec. 9-11: 1157-1162

Barabasi, Albert-Laszlo (2003), *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life,* Penguin Group, Plume: New York

Berk, Richard A. (2008) *Statistical Learning from a Regression Perspective*, Springer: New York

Borsboom, Denny; Mellenbergh, Gideon J. and van Heerden, Jaap (2003) “The Theoretical Status of Latent Variables” *Psychological Review*, Vol. 110, No. 2, 203–219 DOI: 10.1037/0033-295X.110.2.203

Buchanan, M. (2000) *Ubiquity: The Science of History or Why the World is Simpler than We Think*: Phoenix: London

Carson, Richard T., Louviere, Jordan J., Anderson, Donald A., Bunch, David S., Hensher, David A., Johnson, Richard M., Kuhfeld, Warren F., Steinberg, Dan, Swait, Joffre, Timmermans, Harry and Wiley, James B. (1994) “Experimental Analysis of Choice” *Marketing Letters*, Vol. 5, No. 4, Duke Special Issue, Oct., , pp. 351-367

Chatfield, C. (1989) *The Analysis of Time Series: An Introduction*, Chapman & Hall: London

Chatterjee, Sangit, Yilmaz, Mustafa R., Habibuliah, Mohamed and Laudato, Matthew (2000) “An Approximate Entropy Test for Randomness”, *Communications in Statistics - Theory and Methods*, 29: 3, 655 - 675

Conrad, Keith (2011) “Probability Distributions and Maximum Entropy” retrieved 2011 April 23 from http://www.math.uconn.edu/~kconrad/blurbs/analysis/entropypost.pdf

Cook, Matthew (2009) “A Concrete View of Rule 110 Computation” *EPTCS* 1, 2009, pp. 31-55

Cranz, C., and Becker, K., (1921) *Exterior Ballistics: Being a Theoretical Examination of the Motion of the Projectile from the Muzzle to the Target* (Translated from the Second German Edition), His Majesty's Stationery Office: London, (p 383)

Crutchfield, James P. (1994a) “The Calculi of Emergence: Computation, Dynamics, and Induction” *Physica D* special issue on the Proceedings of the Oji International Seminar Complex Systems — from Complex Dynamics to Artificial Reality held 5 - 9 April 1993, Numazu, Japan, SFI 94-03-016

Crutchfield, James P., Ellison, Christopher J. and Mahoney, John R. (2009) “Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information.” *Physical Review Letters* 103, 094101

David, H. A. (1998) “Early Sample Measures of Variability” *Statistical Science*, Vol. 13, No. 4 Nov. pp. 368-377

Denis, A. and Crémoux, F. (2002) “Using the Entropy of Curves to Segment a Time or Spatial Series” *Mathematical Geology*, Vol. 34, No. 8, November: 899-914

Denize, S and Young, L (2007) Concerning Trust and Information, *Industrial Marketing Management*, 36 7 968-982

Dette, Holger; Munk, Axel; and Wagner, Thorsten (1998) “Estimating the Variance in Nonparametric Regression-What is a Reasonable Choice?” *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, Vol. 60, No. 4, pp. 751-764

Dooley, Kevin J. and Van de Ven, Andrew H. (1999) “Explaining Complex Organizational Dynamics” *Organization Science*, Vol. 10, No. 3, Special Issue: Application of Complexity Theory to Organization Science (May - Jun), pp. 358-372

DuPain, Yves; Kamae, Teturo and Mendés France, Michel (1986) “Can One Measure the Temperature of a Curve?” *Archive for Rational Mechanics and Analysis* Volume 94, Number 2, 155-163 Corrigenda, *Arch. Rational Mech. Anal.* 98 (1987) 395

Feldman, David P. and Crutchfield, James P. (1998) “Measures of Statistical Complexity: Why?” *Physics Letters A*, Volume 238, Issues 4-5, 9 February, Pages 244-252

Feldman, David P., McTague, Carl S. and Crutchfield, James P. (2008) “The Organization of Intrinsic Computation: Complexity-Entropy Diagrams and the Diversity of Natural Information Processing” *Chaos* 18, 043106

Frank, Steven A. and Smith, Eric (2010) “A Simple Derivation and Classification of Common Probability Distributions Based on Information Symmetry and Measurement Scale” *Santa Fe Institute Working Papers*, Retrived 2011 March 1 from http://www.santafe.edu/media/workingpapers/10-11-022.pdf

Freund, John E. (1984) *Modern Elementary Statistics*, Sixth edition, Prentice-Hall

Gasser, Theo; Sroka, Lothar; and Jennen-Steinmetz, Christine (1986) “Residual Variance and Residual Pattern in Nonlinear Regression” *Biometrika*, Vol. 73, No. 3 Dec. pp. 625-633

Gell-Mann, Murray (1994a) *The Quark and the Jaguar: Adventures in the Simple and the Complex*, Abacus: London

Gell-Mann, Murray (1995) “What is Complexity?” in *Complexity*, Vol. 1, no. 1 Retrieved 2011 February 20 from http://tuvalu.santafe.edu/~mgm/Site/Publications_files/MGM%20116.pdf

Geweke, John (1992) “Comment: Inference and Prediction in the Presence of Uncertainty and Determinism”, *Statistical Science*, Vol. 7, No. 1 (Feb., 1992), pp. 94-101

Granger, Clive W. J. and Orr, Daniel (1972) ““Infinite Variance” and Research Strategy in Time Series Analysis” *Journal of the American Statistical Association*, Vol. 67, No. 338 Jun. pp. 275-285

Hall, Peter; Kay, J. W.; and Titterington D. M. (1990) “Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression” *Biometrika*, Vol. 77, No. 3 Sep. pp. 521-528

Hall, Peter and Morton, Sally C. (1993) “On the estimation of entropy” *Annals of Institute of Mathematical Statistics* 45, 69–88

Hausser, Jean and Strimmer, Korbinian (2009) “Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks” *Journal of Machine Learning Research* 10: 1469-1484

Holbrook, Morris B. (2003) Adventures in Complexity: An Essay on Dynamic Open Complex Adaptive Systems, Butterfly Effects, Self-Organizing Order, Coevolution, the Ecological Perspective, Fitness Landscapes, Market Spaces, Emergent Beauty at the Edge of Chaos, and All That Jazz, *Academy of Marketing Science Review* Vol 2003. no. 06 (Available: http://www.amsreview.org/articles/holbrook06-2003.pdf)

Johnson, R.A. and Wichern, D.W. (1992) *Applied Multivariate Statistical Analysis*, Prentice-Hall: New Jersey

Kendall, M. G. and Bradford Hill, A. (1953) “The Analysis of Economic Time-Series-Part I: Prices” *Journal of the Royal Statistical Society, Series A (General)*, Vol. 116, No. 1, pp. 11-34

Keogh, Eamonn J. and Pazzani, Michael J. (1998) "An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback" in R. Agrawal, P. Stolorz and G. Piatetsky-Shapiro (eds), *Fourth International Conference on Knowledge Discovery and Data Mining (KDD 1998)*: 239-241

Kvam, Paul H. and Vidakovic, Brani (2007) *Nonparametric Statistics with Applications to Science and Engineering*, John Wiley & Sons, Inc., Hoboken, New Jersey

Langton, Chris G. (1990) "Computation at the Edge of Chaos: Phase Transitions and

Laughlin, Robert M. (2005) *A Different Universe: Reinventing Physics from the Bottom Down*, Basic Books

Liao, T. Warren (2005) "Clustering of Time Series Data—A Survey" *Pattern Recognition* 38, 1857 – 1874

Lloyd, Seth (2000) "Ultimate physical limits to computation" *Nature* (London) 406, 1047–1054

Louviere, Jordan J. (2006) "What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments" *Environmental & Resource Economics* (2006) 34: 173–188

Louviere, J. Street, D. Carson, R. Ainslie, A. Deshazo, J. Cameron, D. Hensher, D. Kohn, R. Marley, T. (2002) "Dissecting the Random Component of Utility" *Marketing Letters* 13:3, 177–193

Mandelbrot, Benoit B. (1963) "The Variation of Certain Speculative Prices" *The Journal of Business*, Vol. 36, No. 4 (Oct), pp. 394-419

Maragos, Petros and Potamianos, Alexandros (1995) "Higher Order Differential Energy Operators" *Signal Processing Letters, IEEE* Volume 2, Issue 8, Aug 1995 Page(s):152 - 154

Marks, Robert E. (1998) "Evolved Perception and Behaviour in Oligopolies" *Journal of Economic Dynamics and Control* 22, 1209-1233

McKelvey, Bill and Andriani, Pierpaolo (2005) "Why Gaussian Statistics are Mostly Wrong for Strategic Organization" *Strategic Organization*, 3(2): 219–228

Mendenhall, William; Scheaffer, Richard L.; and Wackerly, Dennis D. (1981) *Mathematical Statistics with Applications*, Second edition, Duxbury Press; Boston, Massachusetts

Miller, John H. and Page, Scott E. (2004) "The Standing Ovation Problem", *Complexity*, 9: 8-16

Milne-Thomson, L. M. (1933) *The Calculus of Finite Differences*, Macmillan: London

Mitchell, Melanie; Crutchfield, James P. and Hraber, Peter T. (1999) “Dynamics, Computation, and the “Edge of Chaos”: a Re-Examination” In *Complexity: Metaphors, Models, and Reality*, G. A. Cowan, D. Pines, and D. Meltzer, Eds. Perseus Books Advanced Book Classic Series. Perseus Books, Cambridge, MA, 497-513.

Miśkiewicz, Janusz (2010) “Entropy Correlation Distance Method. The Euro Introduction Effect on the Consumer Price Index” *Physica A* 389 1677-1687

Möller-Levet, C. S., Cho, K.-H., Wolkenhauer, O. (2003) “Microarray Data Clustering Based on Temporal Variation: FCV with TSD Preclustering”, *Appl. Bioinformatics* 2 (1) 35–45

Möller-Levet, C. S., Klawonn, F., Cho, K.-H. and Wolkenhauer, O. (2003) “Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points”, *LNCS, Proceedings of the IDA2003*

Möller-Levet, C. S., Cho, K.-H., Yina, H. and Wolkenhauer, O. (2003) *Clustering of Gene Expression Time-Series Data*, Technical Report 11-2003, SBI, Department of Computer Science, University of Rostock, Germany

Möller-Levet, C. S., Klawonn, F., Cho, K.-H., Yina, H. and Wolkenhauer, O. (2005) “Clustering of Unevenly Sampled Gene Expression Time-Series Data” *Fuzzy Sets and Systems* 152 49–66

Olson, David L. and Delen, Dursun (2008) *Advanced Data Mining Techniques*, Springer-Verlag: Berlin, Heidelberg

Pincus, Steve. M. (1991) “Approximate entropy as a measure of system complexity” *Proc. Natl. Acad. Sci*. USA 88:2297-2301

Pincus, Steve M. (2008) “Approximate Entropy as an Irregularity Measure for Financial Data”, *Econometric Reviews*, 27:(4-6): 329-362

Ramos, O., Altshuler, E. and Maloy K. J. (2009) ‘Avalanche Prediction in a Self-Organized Pile of Beads’ *Physical Review Letters*, PRL 102, 078701 (2009)

Rand, W. & Rust, R. T. 2011. Agent-based modeling in marketing: Guidelines for Rigor. *International Journal of Research in Marketing*, In Press, Accepted Manuscript.

Rissanen, Jorma (2007) *Information and Complexity in Statistical Modeling*, Springer: New York

Sabbione, Juan I. and Velis, Danilo (2010) “Automatic First-Breaks Picking: New Strategies and Algorithms” *Geophysics*, Vol. 75, No. 4 July-August; p. V67–V76, 10.1190/1.3463703

Schelter, Björn; Winterhalder, Matthias; Timmer, Jens; eds (2007) *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, Wiley-VCH

Shalizi, Cosma R. (2009) “Estimating Distributions and Densities” retrieved 2011 April 26, from http://www.stat.cmu.edu/~cshalizi/350/lectures/28/lecture-28.pdf

Shannon, C.E. (1948) “A Mathematical Theory of Communication” *Bell System Technical Journal* Vol. 27, July, October, pp. 379-423, 623-656

Smith AE, Humphreys MS. (2006) Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavioural Research Methods*; 38 (2): 262-279

Soofi, Ehsan S. and Retzer, J.J. (2002) “Information Indices: Unification and Applications” *Journal of Econometrics* 107:17 - 40

Stanley, M. H. R., Amaral, L. A. N., Gopikrishnan, P. and Plerou, V. Plerou (2000) “Scale Invariance and Universality of Economic Fluctuations” *Physica A* 283 31–41

Sutner, Klaus (2010) “Cellular Automata, Decidability and Phasespace”, *Fundamenta Informaticae* 104 (2010) 1–20, DOI 10.3233/FI-2010-300

Todorovski, Ljupco; Cestnik, Bojan; Kline, Mihael; Lavrac; Dzeroski, Saso (2002) “Qualitative Clustering of Short Time Series: A Case Study of Firms Reputation Data” *Conference on Data Mining and Warehouses (SIKDD 2002)*, Ljubljana, Slovenia

Victor, Jonathan D. (2002) “Binless Strategies for Estimation of Information from Neural Data” *Physical Review E*, 66, 051903

von Neumann, J.; Kent, R. H.; Bellinson, H. R.; Hart B. I. (1941) “The Mean Square Successive Difference” *The Annals of Mathematical Statistics*, Vol. 12, No. 2 (Jun), pp. 153-162

Watts, Duncan J. and Strogatz, Steven (1998) “Collective Dynamics of “Small-World” Networks” *Nature*, 393:440-442

Wang, X., Smith, K., and Hyndman R. (2006) “Characteristic-Based Clustering for Time Series Data” *Data Mining and Knowledge Discovery*, 13(3):335–364

Weisberg, Sanford (1985) *Applied Linear Regression,* Second Edition, John Wiley & Sons, Inc.

Weisberg, Sanford (2005) *Applied Linear Regression,* Third Edition, John Wiley & Sons, Inc.: Hoboken, New Jersey

Weiss, Robert E. (2005) Modeling Longitudinal Data, Springer Science+Business Media: New York

Weisstein, Eric W. (2010) “Finite Difference” *MathWorld*: A Wolfram Web Resource. Retrieved 2010 March 7, http://mathworld.wolfram.com/FiniteDifference.html

Weisstein, Eric W. (2010b) “Divided Difference” *MathWorld*: A Wolfram Web Resource. Retrieved 2010 March 7 http://mathworld.wolfram.com/DividedDifference.html

Wiesner, Karoline; Gu, Mile; Rieper, Elisabeth and Vedral, Vlatko (2009) "Information Erasure Lurking Behind Measures of Complexity" *arXiv:0905.2918v1 [physics.data-an]*

Wilkinson, Ian and Young, Young (2002) "On Cooperating: Firms, Relationships and Networks" Journal of Business Research 55 (2), 123-133,

Wilkinson, Ian and Young, Louise (2005) "Toward a Normative Theory of Normative Marketing Theory" Marketing Theory, 5 4 363-396

Wilkinson, Ian and Young, Louise (2011) Toward a Business Network Simulation System for Researchers, Educators, Managers and Policymakers, Business Marketing Management Conference, Tampere,, Finland, May.

Wilkinson, Ian; Young, Louise; Marks, Robert; Bossomaier, Terry and Held, Fabian (2011) Toward a Business Network Agent-Based Modeling System, International Conference on Complex Systems, Boston, June.

Wolfram, Stephen (2002) *A New Kind of Science*, Wolfram Media, Inc: Champaign, Il

Wolfram, Stephen (2006) "The State of NKS in 2006" keynote at the *NKS 2006 Conference* in Washington D.C. Retrieved from http://www.stephenwolfram.com/publications/recent/nks2006/ on 2010 April 12

Wong, Charles; Wilkinson, Ian and Young, Louise (2010) Towards an Empirically Based Taxonomy of Buyer-Seller Relations in Business Markets, Journal of the Academy of Marketing Science. 38 6 720–737 (also available at http://www.springerlink.com/content/e8g944664u2t5506/

Wu, Hulin and Zhang, Jin-Ting (2006) *Nonparametric Regression Methods for Longitudinal Data Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey

Wuensche, A. (1998) "Discrete dynamical networks and their attractor basins" In R. Standish, B. Henry, S. Watt, R. Marks, R. Stocker, D. G. Green, S. Keen, & T. Bossomaier (Eds.), *Complex systems '98* (pp. 3-22). Complexity Online Network

Yitzhaki, Shlomo (2003) "Gini's Mean difference: a superior measure of variability for non-normal distributions" *Merton - International Journal of Statistics*, vol. LXI, n. 2, pp. 285-316