emailkbh904@alumni.ku.dk

Statistics, Probability and Chaos – a Review

Adrián A. Sousa-Poza
Faculty of Science, University of Copenhagen
(Dated: March 8, 2023)

Contents

I. Introduction

Chaos refers to the unpredictable behaviour of a system that appears to be highly sensitive to its initial conditions [1]. A small change in the starting conditions can lead to vastly different outcomes or trajectories. This makes using traditional statistical methods for chaotic systems very challenging as small differences in initial conditions can lead to very different outcomes. Thus, estimating a probability distributions of the underlying system variables can become quite problematic. Especially when considering time series, which refer to a sequence of data points that are measured at regular time intervals [2], can be difficult to analyse as over time the behaviour of the underlying system becomes untraceable and therefore unpredictable. A well-known example is the study of weather and climate. While we may be able to predict the weather for tomorrow, it is impossible to make accurate predictions for tomorrow next year. Edward Lorenz (1917 – 2008), an American mathematician and meteorologist, famously used the example of, given a large enough time frame, a butterfly causing a tornado in Brazil by flapping it's wings in the rain forests of Brazil [3]. Berliner [4] addresses in his paper *Statistics, Probability and Chaos (1992)* how statisticians can deal with chaos and how they can contribute to the analysis of systems showing chaotic behaviour.

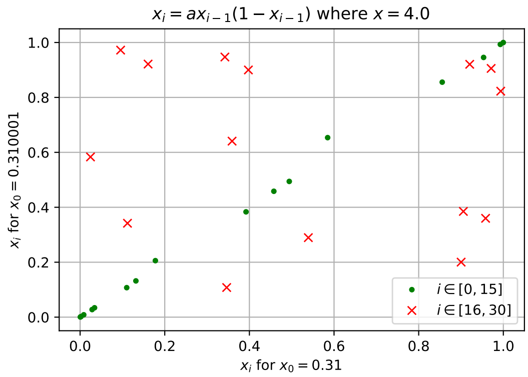

FIG. 1: Scatter plot of the time series $x_i = ax_{i-1}(1 - x_{i-1})$ with $x = 4.0$ and the initial conditions $x_0 = 0.31$ and $x_0 = 0.310001$.

II. Review

A. Mathematics, Probability and Chaos

A simple way to emphasise the importance of analysing the chaotic nature of a system is by studying the time series $x_i = ax_{i-1}(1 - x_{i-1})$ (logistic map) with $x = 4.0$ [4]. Creating a series scatter plot of the initial conditions $x_0 = 0.31$ and $x_0 = 0.310001$ (see figure 1), we can see that while the first fifteen points of the series with two slightly different initial conditions seem to correlate quite strongly, the last fifteen points of the two time series are extremely different from each other.

Berliner [4] emphasises that there are no universally accepted definitions of chaos, but it can be described in terms of the following definition.

Definition 1 (sensitivity to initial conditions). *$f : D \to D$ displays sensitivity to initial conditions if there exists $\delta > 0$ such that for any x in D and any neighbourhood V of x, there exists a y in V and $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.*

This definition defines sensitivity of initial conditions as the existence of points arbitrarily close to x that will eventually separate from x during the evolution of the system over time. It is important to point out that this is not the only definition of sensitivity or chaos. However, all definitions try to emphasise that given a chaotic system (or a system sensitive to initial conditions) small changes in initial conditions will lead to very different outcomes.

Furthermore, Berliner [4] discusses the interrelations between chaos and randomness, uncertainties, and ergodic theory. Having uncertainty in a complex system will inevitably lead to the use of random or probabilistic methods. It is argued that if even small uncertainties exist in a system which shows high sensitivity to initial conditions, we will end up analysing a chaotic system. Therefore, uncertainties make a system probabilistic. As an example, we can take the logistic map, where if we have no knowledge about x_0, we will end up analysing a non-deterministic system. Under certain conditions this also leads to chaos being random. Note that in this sense randomness is defined as a sequence of events lacking any discernible pattern or predictability. Berliner [4] mentions that through the consideration of ergodic theory a strong relationship between deterministic chaos and randomness can be found. Ergodic theory seeks to answer the question of when we can expect the average of data in one experiment to be the same as the average of the data over similar replicates at a fixed time. Similarly Wikipedia defines ergodicity as *the idea that a point of a moving system, either a dynamical system or a stochastic process, will eventually visit all parts of the space that the system moves in, in a uniform and random sense.*

III. Data Analysis and Chaos

In order to analyse chaos there are several methods we can make use of. A very common measure are Lyapunov exponents. They quantify the degree of sensitivity of a system to small changes in initial conditions. The concept is based on the idea that two trajectories in phase space that start out very close to each other will eventually diverge from each other (note the similarity to definition 1) at an exponential rate, with the rate of divergence being given by the Lyapunov exponent.

Berliner [4] discusses the challenge of identifying and analysing underlying determinism in chaotic data. The example presented involves a data set of computer-generated iterations of the logistic map with $a = 4$ and a randomly chosen initial conditions. A simple time series plot of the data would suggest that the data is random, but fitting an model and examining a scatterplot of x_{i-1} versus x_i would reveal the deterministic structure of the logistic map. Searching for structure in chaotic systems like the logistic map can for instance be done with Poincaré maps, which are a mathematical tool used to analyse the dynamics of a continuous-time dynamical system by examining its behaviour at discrete intervals [1].

Finally, Berliner [4] discusses how statistics can be used to analyse chaotic systems. One approach is parametric statistical analysis, which involves specifying a dynamical function driving the system and estimating its unknown parameters based on observed data with error. However, likelihood functions based on such models can be extremely complex and difficult to work with, and initial conditions can greatly affect predictions. A way to account for that is working with Bayesian statistics, which can be used to incorporate uncertainties in initial conditions to create forecasts. Nevertheless, long-term predictions of chaotic processes are almost impossible.

IV. Conclusion

In conclusion, the analysis of chaotic systems presents a challenge for statisticians as small differences in initial conditions can lead to vastly different outcomes, making it difficult to estimate the probability distributions of system variables. The sensitivity to initial conditions of chaotic systems is what makes them so unpredictable and challenging to analyse. A simple example of the logistic map demonstrates how small differences in initial conditions can lead to vastly different outcomes. However, despite the challenges, statisticians can contribute to the analysis of chaotic systems by using methods such as Lyapunov exponents, Poincaré plots, maximum likelihood estimation and Bayesian inference. As many fields such as meteorology, biology, physics, and economics deal with chaotic systems, it is important to be familiar with the concept of chaos in order to understand how a system which is sensitive to initial conditions can affect the underlying data.

[1] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (CRC Press, 2018).

[2] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications* (Springer International Publishing, 2017).

[3] C. Rouvas-Nicolis and G. Nicolis, Scholarpedia **4**, 1720 (2009).

[4] L. M. Berliner, Statistical Science **7**, 10.1214/ss/1177011444 (1992).