

Lecture 9: Theoretical approaches to predictability

November 9, 2014

1 Small error case

The traditional theoretical approach to predictability pioneered in the late 1960s and 1970s focusses on the growth of infinitesimal errors within a dynamical system. As we saw in the previous lecture this is usually the conceptual approach taken in practical studies as well since they often rely effectively on a linearization assumption. Consider a general dynamical system of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \quad (1)$$

where the bold quantities are vectors (usually assumed finite dimensional). If we linearize this about a particular set of initial conditions then the evolution equation for small perturbations (errors) is approximately given by the linear equations

$$\frac{d(\delta\mathbf{x})}{dt} = \nabla\mathbf{F}|_{\mathbf{x}=\mathbf{x}(0)} \bullet \delta\mathbf{x}$$

Under rather general assumptions regarding the dynamical system given by (1) it was shown by [9] that these perturbations can almost always be written for large t in the form

$$\delta x^i(t) = c_j e_j^i \exp(\lambda_j t) \quad (2)$$

where we are assuming the summation convention. Here the vectors \mathbf{e}_j are referred to as Lyapunov vectors while the real λ_j are called characteristic Lyapunov exponents which are normally assumed ordered as a decreasing set by the index j . The \mathbf{e}_j are an orthonormal basis which varies according to the initial conditions $\mathbf{x}(0)$ selected. The exponents λ_j on the other hand **are independent of initial conditions** which is the real content of this result. If we suppose that our initial condition error is a small sphere within the dynamical system space then equation (2) shows that given sufficient time this evolves into an ellipse with principal axes directions given by the Lyapunov vectors and with length given by $e^{\lambda_j t}$. Note that the Lyapunov vectors are the time asymptotic singular vectors and the Lyapunov exponents are functionally related to the corresponding time asymptotic singular values.

Given the ordering and assuming that some of the exponents are positive (an assumption that the system is chaotic) then with sufficient time this ellipse and hence errors will be dominated by the direction of the first or leading Lyapunov vector. For this reason the leading Lyapunov exponent is often an object of considerable focus in chaotic dynamical systems as it indicates an approximate (and general) e-folding time for small errors. It is then assumed that this indicates a time scale on which errors become large and hence predictability is lost. Of course this is rather imprecise as we will see later in detail since it only applies to small errors but nevertheless it is a rough indicator of a predictability limit timescale within a system. As an interesting example it may be calculated in the solar system and the estimated reciprocal is between 10^6 and 10^7 years which suggests that predictions of eclipses and the like will be very accurate for a considerable period into the future. Given a period comparable with geological time scales however the system appears unpredictable which is confirmed by numerical experiments. The chaotic behavior originates because of many body interactions within the system and a two body system does not have this property.

The Lyapunov exponents have played a central role in the study of chaotic dynamical systems of low dimension. In general the rigorous results obtained have been fairly restrictive partly due to the fact that the exponents are difficult to calculate except numerically. One important conjecture due to Kaplan and Yorke (see [7]) asserts that the fractal dimension of the attractor (i.e. invariant measure) of a dynamical system is given by

$$D_{KY} \equiv J + \frac{\sum_{i=1}^J \lambda_i}{|\lambda_{J+1}|}$$

where J is the smallest integer such that $\sum_i^J \lambda_i \geq 0$. This has been proven for hyperbolic maps in the plane by Courant faculty member L.S. Young (see [11]). It has been verified to a reasonable degree of accuracy in many low dimensional systems using numerical methods. D_{KY} has important practical application since given the output of many dynamical systems, it is typically more easily calculated than other fractal dimension measures.

The positive exponents, which give the error growth (not decay) rates, may also be related rigorously to the so-called Kolmogorov-Sinai entropy. This is defined as follows: Consider a finite partitioning Λ of the state space of the dynamical system. At any time the dynamical system clearly lies within one of the partition elements. If we suppose that the system is in equilibrium so that probabilities are given by an invariant measure then one can define a joint probability of the partitioned system at a sequence of N equally spaced time values with associated random variables X_1, X_2, \dots, X_N each of which has outcomes labelling the partitions. This clearly forms a stochastic process of the type studied in Lecture 3. The joint entropy of this time invariant sequence of random variables is clearly $H^\Lambda(X_1, X_2, \dots, X_N)$ where the superscript reminds us that it depends on the partition chosen. The usual modified entropy rate

(per unit time) is then using the entropy chain rule

$$h(\Lambda, \tau) = \lim_{N \rightarrow \infty} \frac{1}{\tau} (H^\Lambda(X_{N+1}, X_N, \dots, X_1) - H^\Lambda(X_N, \dots, X_1))$$

where τ is the time interval. Within certain general classes of deterministic systems it is possible to show that this entropy rate converges as the time interval τ decreases (see [2]). Interestingly in stochastic systems such as some of those considered in Lecture 5 it does not (see [6]) and then other entropies are relevant. Naturally of course it usually does depend on the partitioning Λ chosen however if one considers all possible partitions then the supremum of this limiting modified rate defines the Kolmogorov-Sinai entropy:

$$h_{KS} \equiv \lim_{\tau \rightarrow 0} \sup_{\Lambda} (h(\Lambda))$$

Because of the supremum and limiting definition h_{KS} is best considered as a fine grained quantity of a dynamical system. A very important result in dynamical system theory was proven in the mid 1970s by Pesin (see [10]) relating this to the positive Lyupanov exponents of the system:

$$h_{KS} \leq \sum_{\lambda_i > 0} \lambda_i$$

This connection has been studied at length and a rigorous review of results including when equality is achieved in the last equation can be found in [5]. Given the fact as noted above that the Lyupanov exponents are functionally related to the growth/decay of small volumes and the differential entropy evolution of Lecture 5 is related to the expectation of volume changes within the dynamical system one may wonder about the connection between the two entropies. Note though that that the Kolmogorov-Sinai entropy is related to the positive growth directions only and refers to a system in equilibrium whereas the differential entropy evolution considered in Lecture 5 is connected to systems not in equilibrium and is affected by negative volume growth directions as well as the positive. An interesting (but possibly difficult!) student project would however be to explore the connections here in depth. A good starting point is the review article [6].

2 Finite error case

The limitation of the previous section is that it applies only to the infinitesimally small error case. Thus use of the largest Lyupanov exponent to define a predictability limit tacitly assumes that for the long times appropriate to discussion of such a limit that equation (2) holds which is of course very questionable. As errors grow to a size comparable with typical equilibrium fluctuations this equation is no longer valid. This situation has led to the generalization of the approach of the previous section to finite perturbation analysis and associated

generalized Lyapunov vectors and exponents (see [1] and [3]). Associated with these new exponents is a “coarse grained” version of h_{KS} which is defined on a finite “coarse” partition of the dynamical system space (see [6]). This area has seen considerable progress in the last 20 years and a good reference is the recent book [4].

3 Information theory framework

The small error limitations of the approaches outlined above led the lecturer and co-workers to a different approach 10 years ago rooted more directly in information theory and in principle of a rather more general character:

There are two probability distributions of importance to predictability. The first is the *prediction* distribution which is the probability distribution for the random variable which we wish to predict at a particular time. In general, as we saw in the previous lecture, we assume that at the initial time that this is specified due to the nature of the observation network. A dynamical model of some kind is used then to evolve this specified distribution forward in time. We refer to this process as “statistical prediction”. For very large times we shall assume that the prediction distribution converges asymptotically to the second important kind of distribution namely the *equilibrium* distribution which we further assume is unique¹. If the dynamical system is ergodic (a common hypothesis for realistic systems) then this latter distribution will also be the *historic* or *climatological* distribution. Note that the equilibrium distribution may not necessarily be time invariant however we shall assume that the dynamical system under consideration is subject only to external *periodic* forcing and so the equilibrium distribution is also periodic. The earth system closely approximates such a system with the dominant external forcings being the annual and diurnal cycles caused by the Earth’s rotation about the sun and it’s axis.

In general if one knows nothing concerning the initial conditions of a dynamical system then the best assumption concerning relevant random variables is that they have the equilibrium or historical distribution. Statistical prediction is then the process of using initial condition information to modify this distribution. In a perfect prediction one would modify it to be a delta function about a particular value.

Clearly this process is analogous to the Bayesian paradigm for learning discussed in the second lecture. The prior (“before enlightenment”) in this case is the equilibrium distribution while the posterior (“after enlightenment”) is the prediction distribution. Motivated by this analogy, the relative entropy of the two distributions was introduced in [8] as a measure of the importance or utility of the statistical prediction process.

In addition to it’s interpretation as the utility of the process of statistical prediction, relative entropy also measures the asymptotic equilibration process

¹These assumptions are based on empirical observation of the behaviour of numerical models rather than on rigorous results. The complexity of realistic dynamical systems generally forbids the latter.

of the prediction distribution as it converges toward the equilibrium distribution. Recall that it is used precisely in this context when analyzing the Fokker Planck equation of stochastic differential equations. When the relative entropy falls to a value close to zero then one can conclude that initial condition data is basically useless for a prediction. Another way of stating this is to say that the measure of predictability proposed is also a measure of the degree of statistical disequilibrium of the dynamical system under study.

Finally as a measure of predictability, relative entropy has three desirable properties as we saw in Lecture 5. Firstly it is invariant under a general well behaved non-linear transformation of state variables. Secondly it has a clear interpretation in terms of the coarse graining of state space in that the differential form is the natural limit of the coarse grained version. Thirdly in stochastic Markov processes the relative entropy obeys a monotonicity property in time i.e. the equilibration process, as measured by the relative entropy, never reverses and in many important cases strictly declines. As an information theoretic predictability functional the relative entropy appears unique in this respect although this is unproven as far as the lecturer is aware.

It is important to stress that the interpretation of relative entropy given above is valid under the assumption that the model is perfect i.e. not subject to model error. Naturally this assumption is approximately true only in a realistic situation. The point is that one is able to analyze utility drops due only to initial condition error growth.

References

- [1] E. Aurell, G. Boffetta, A. Crisanti, G. Paladin, and A. Vulpiani. Growth of noninfinitesimal perturbations in turbulence. *Phys. Rev. Lett.*, 77(7):1262–1265, 1996.
- [2] P. Billingsley. Ergodic theory and information. 1965.
- [3] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani. Predictability: a way to characterize complexity. *Phys. Rep.*, 356(6):367–474, 2002.
- [4] P. Castiglione, M. Falcioni, A. Lesne, and A. Vulpiani. *Chaos and coarse graining in statistical mechanics*. Cambridge University Press New York, NY, USA, 2008.
- [5] J.P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57(3):617–656, 1985.
- [6] P. Gaspard and X.J. Wang. Noise, chaos, and (ε, τ) -entropy per unit time. *Phys. Rep.*, 235(6):292–343, 1993.
- [7] J. Kaplan and J. Yorke. Chaotic behavior of multidimensional difference equations. *Functional Differential equations and approximation of fixed points*, pages 204–227, 1979.

- [8] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, 59:2057–2072, 2002.
- [9] VI Oseledec. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *English transl. Trans. Moscow Math. Soc.*, 19:197–221, 1968.
- [10] Y. Pesin. Lyapunov characteristic exponents and ergodic properties of smooth dynamical systems with an invariant measure. In *Sov. Math. Dokl.*, volume 17, pages 196–199, 1976.
- [11] L.S. Young. Dimension, entropy and lyapunov exponents. *Ergodic Theory Dynam. Systems*, 2(1):109–124, 1982.