

On the probability of chaos in large dynamical systems: A Monte Carlo study

W. Davis Dechert^{a,*}, Julien C. Sprott^b, David J. Albers^b

^aDepartment of Economics, University of Houston, Houston, USA

^bDepartment of Physics, University of Wisconsin, Madison, USA

Received 29 August 1995; accepted 1 July 1997

Abstract

In this paper we report the result of a Monte Carlo study on the probability of chaos in large dynamical systems. We use neural networks as the basis functions for the system dynamics and choose parameter values for the networks randomly. Our results show that as the dimension of the system and the complexity of the network increase, the probability of chaotic dynamics increases to 100%. Since neural networks are dense in the set of dynamical systems, our conclusion is that most large systems are chaotic. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Lyapunov exponents; Frequency of chaos; Neural networks

1. Introduction

In Brock (1993) there is an argument that the larger the dimension of a non-linear dynamical system, the larger the probability that the system dynamics have a positive Lyapunov exponent. (Clearly this argument can only hold in a probabilistically generic sense.) From this the conclusion is drawn that "... it is not absurd to expect that the chances are high for obtaining a positive Lyapunov exponent for a dynamical system on R^n 'drawn at random' ..." when the dimension, n , is large. In this paper we use Monte Carlo methods to investigate to what extent this is true.

* Correspondence address: Department of Economics, Social Systems Research Institute, University of Wisconsin-Madison, 1180 Observatory Drive, 6470 Social Science Building, Madison, WI 53706-1393, USA. Tel.: + 1 608 262 0446; fax: + 1 608 263 3876; e-mail: ssri@facstaff.wisc.edu

We use neural networks to represent dynamical systems. Sprott (1993) used polynomial functions to represent dynamical systems and found that as the complexity¹ of the system increased, the fraction of the bounded systems that are chaotic increased for continuous systems (ordinary differential equations) and decreased for discrete systems (maps). An advantage of using neural networks is that the dynamical system is always bounded. Since we are not testing specific dynamical systems, it is important to choose a class of representative functions that are dense in the space of all dynamical systems. Neural networks (like polynomials) are indeed dense in the set of continuous functions on a bounded interval. These are the only properties of neural nets that we use, and any other class of representative functions with these properties could also be used for this type of analysis. One issue that we do not pursue in this paper is the question of the robustness of our findings with respect to the specification (i.e., polynomials, neural networks, etc.) of the set of test functions.

In this paper we report results on discrete time dynamical systems with time-delay coordinates. There is an argument by Takens (1981) that shows that these maps are generic in the space of dynamical systems. We will make this notion more precise in Section 3.

2. Chaotic dynamics

In this paper we report the result of a Monte Carlo study of the likelihood that dynamical systems exhibit chaotic solutions.² We carried out this investigation using single layer, feedforward neural networks. Since a neural network is parametrically defined, it is a simple matter to generate the parameters at random and to count the occurrences of chaos. There are three scaling variables that we used: the number of neurons in the single hidden layer, the dimensionality of the system, and the spread in the distribution of the randomly chosen weight parameters. Our findings are that as the dimension of the system increases, the probability of chaos approaches 1.

Hornik et al. (1989), Hornik et al. (1990) showed that any continuous function can be uniformly approximated (on compact sets) by a neural network. Thus, they can be used as proxies for the equations that describe the dynamics of a system. Since our results show that the more complex the system, the greater the probability of chaos, we surmise that the probability of chaotic dynamics increases as the complexity of the system increases.

¹ By complexity of a dynamical system we are referring to the dimension of the system and the degree of the polynomial used to describe the dynamics of the system.

² For the purpose of this numerical study, we have chosen the definition of chaos to be that the largest Lyapunov exponent of the system is positive.

3. Feedforward networks

The single-layer feedforward networks that we use in this paper are

$$f(y) = \sum_{i=1}^n \beta_i \phi \left(s w_{i0} + s \sum_{j=1}^d w_{ij} y_j \right), \tag{1}$$

where $f: \mathbf{R}^d \rightarrow \mathbf{R}$. The parameters β_i and w_{ij} are real, and s is the scaling factor on the weights. The function ϕ is called an *activation unit* or a *neuron*, n is the number of neurons and d is the dimension of the system. The function ϕ , also called a *squashing function*, is typically taken to be the logistic function

$$\phi(u) = \frac{1}{1 + e^{-u}}$$

or the hyperbolic tangent function

$$\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}.$$

From a theoretical point of view, these two representations are equivalent since a simple relationship exists between them:

$$\tanh(u) = 2\phi(2u) - 1.$$

We will denote the class of neural networks (functions of the form in Eq. (1) as the β_i and w_{ij} range over all possible real values) with n neurons and dimension d by $\Sigma_{n,d}$ and the collection of all neural networks by

$$\Sigma = \bigcup_{n,d \geq 1} \Sigma_{n,d}.$$

For our purposes we shall assume that there is enough smoothness present so that the system dynamics are representable by an element of $C^2(M, M)$, the set of twice continuously differentiable functions from a compact manifold, M , of dimension m into itself. These dynamical systems are of the form

$$x_{t+1} = F(x_t), \tag{2}$$

where $F: M \rightarrow M$ and $x_t \in M$. One way to form a parametric class of functions that approximates these systems is to use a neural network for each of the component functions of the system.³ For computational reasons we chose a different approach and used a single neural network to generate a ‘time series’ of scalar data. A single network, $f \in \Sigma_{n,d}$, defines a dynamical system on \mathbf{R}^d by

$$y_t = f(y_{t-d}, y_{t-d+1}, \dots, y_{t-1}), \tag{3}$$

³ This was the approach of Doyon et al. (1993) and the references cited therein.

where $y_t \in R$. There is a connection between the two types of representations of dynamical systems in Eqs. (2) and (3). Systems described by Eq. (3) are easily written in the form of Eq. (2) via the mapping of R^d to itself:

$$(y_1, y_2, \dots, y_d) \rightarrow (y_2, y_3, \dots, f(y_1, y_2, \dots, y_d)).$$

Thus they form a subset of the d -dimensional dynamical systems. How about those systems that are not of the form Eq. (3)? In Takens (1981) is shown that systems of the form (2) which are diffeomorphisms embed (generically) in R^d for some $d \leq 2m + 1$. Thus, there is an open and dense set of dynamical systems, each element of which is topologically conjugate to a system of the form (3). These latter systems can be uniformly approximated (on compacta) by neural networks.

4. Numerical results

In this section we present some of the numerical results from a Monte Carlo study on the probability of chaos in dynamical systems.⁴ For the system dynamics we used Eq. (1) with the hyperbolic tangent function:

$$y_t = \sum_{i=1}^n \beta_i \tanh\left(s w_{i0} + s \sum_{j=1}^d w_{ij} y_{t-j} \right),$$

where the matrix of weights, w , and the vector of coefficients, β , were generated by a pseudo-random number generator.⁵ The elements of w are iid normal with mean zero and unit variance. There is a scaling parameter, s , which can also be interpreted as the standard deviation of the weights. The elements of β are

$$\beta_i = \frac{u_i}{\sum_{j=1}^n u_j}$$

where the $\{u_i\}$ are iid uniform on $[0, 1]$. Note that the coefficients are scaled so that $\sum_{i=1}^n \beta_i = 1$. With this scaling, $f: [-1, 1]^d \rightarrow [-1, 1]$. This scaling is completely general. If without the scaling $\sum_{i=1}^n \beta_i = r$, then $f: [-r, r]^d \rightarrow [-r, r]$. By scaling the inputs y by $1/r$, the weights w by r and the vector β by $1/r$, the function now maps $[-1, 1]^d$ to $[-1, 1]$, and the resulting scaled dynamics are identical, up to the scaling factor. The Lyapunov exponents of the scaled system are the same as those of the unscaled system.

⁴ Each of the authors independently wrote a program to calculate the largest Lyapunov exponent for the neural network system. The programs and data are available at the World Wide Web site: <http://sprott.physics.wisc.edu/neural/>.

⁵ The pseudo random numbers were generated by the algorithm of l'Écuyer (1988) as programmed in Press et al. (1992).

The largest Lyapunov exponent was calculated numerically by randomly selecting an initial point (y_0, \dots, y_{d-1}) and a nearby point (v_0, \dots, v_{d-1}) . Define

$$\Delta y_t = (y_t - v_t, \dots, y_{t+d-1} - v_{t+d-1})$$

and let $|\Delta y_0| = \varepsilon$. Both points were advanced by one time step and the scaling ratio of

$$\frac{|\Delta y_1|}{|\Delta y_0|}$$

was recorded. The vector Δy_1 is then rescaled to length ε , and the new neighbor,

$$(v_1, \dots, v_d) = (y_1, \dots, y_d) + \Delta y_1$$

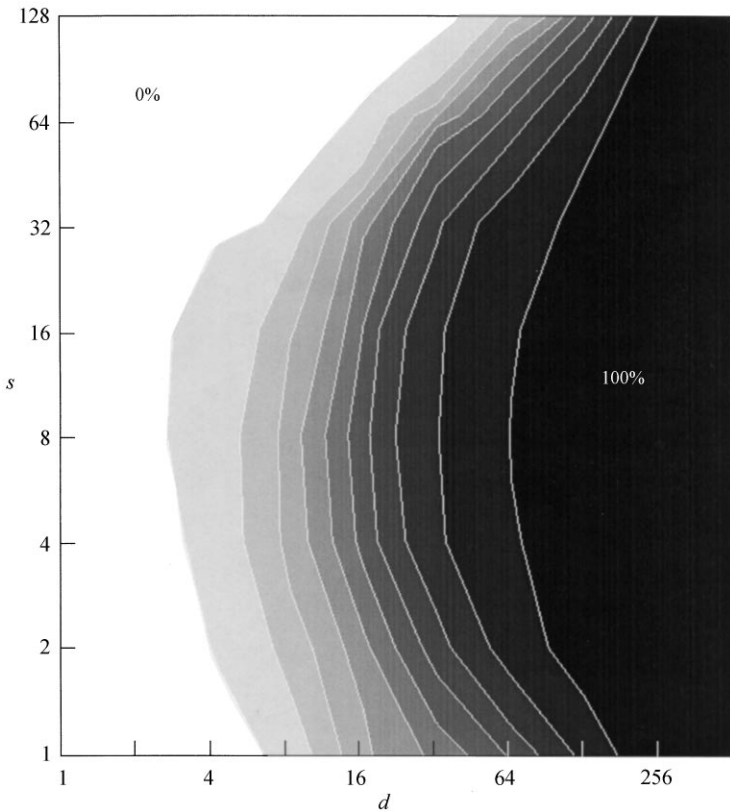


Fig. 1. Contour plot for $n = 8$ with increasing d and s showing that the probability of chaos approaches unity as d becomes large.

was advanced one time period along with (y_1, \dots, y_d) . This process is then repeated, and the largest Lyapunov exponent was estimated by the average of the logs of the scalings:

$$\lambda = t^{-1} \sum_{s=0}^{t-1} \ln \left(\frac{|\Delta y_{s+1}|}{|\Delta y_s|} \right),$$

where t is the number of iterations of the map.

Figs. 1 and 2 show the percentage of chaotic systems (positive Lyapunov exponent) for a collection of approximately 8 500 000 single-layer feedforward neural networks. In Fig. 1 the number of neurons is constant at $n = 8$, and in Fig. 2 the spread is constant at $s = 8$. Note that as the dimension increases, the probability of chaos increases to 1. Second, note the ‘C’ shape of the contour

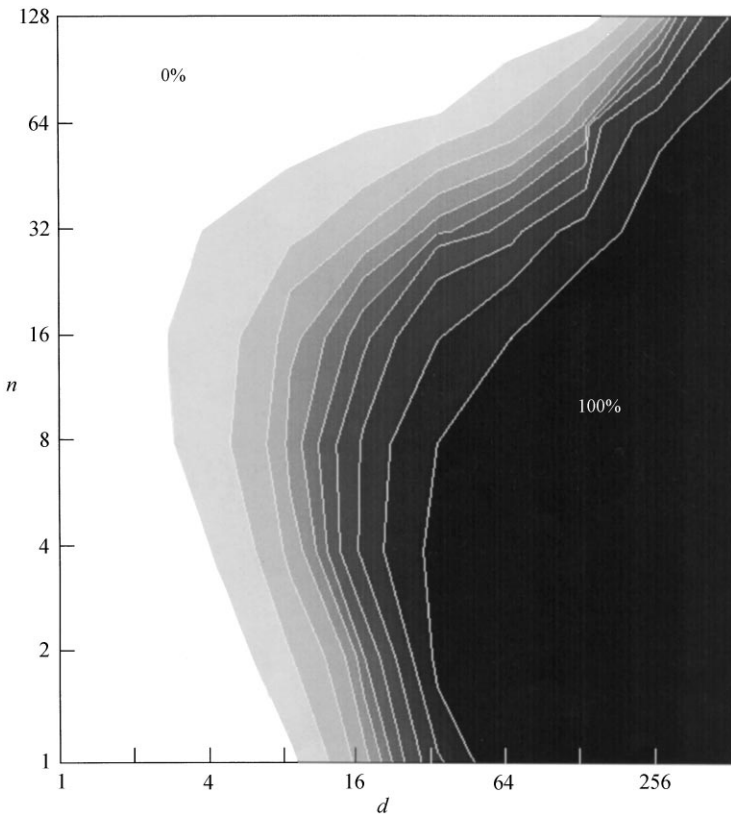


Fig. 2. Contour plot for $s = 8$ with increasing n and d showing that the probability of chaos approaches unity as d becomes large.

lines. The shape in figure one can be explained in terms of the activation function,

$$\begin{aligned} \tanh(u) &= \frac{e^u - e^{-u}}{e^u + e^{-u}} \\ &= \frac{e^{2u} - 1}{e^{2u} + 1}. \end{aligned}$$

When $u \approx 0$ then $\tanh(u) \approx u$ and the function is nearly linear. When $|u| \gg 1$ then $|\tanh(u)| \approx 1$ and the function is nearly constant. For $u = sw_{i0} + s\sum_{j=1}^d w_{ij}y_j$ the conditional expectation of u given y_1, \dots, y_d is

$$E[u|y_1, \dots, y_d] = 0,$$

and the conditional variance of u given y_1, \dots, y_d is

$$E[u^2|y_1, \dots, y_d] = s^2 \left(1 + \sum_{j=1}^d y_j^2 \right).$$

For a given value of the dimension parameter, d , the tanh function will be nearly linear for small values of the spread, s , and will be constant at ± 1 for large values of the spread. Linear systems cannot exhibit chaotic behavior, and when the activation functions are all ± 1 , the number of states is finite and the output is periodic, which also cannot be chaotic.⁶

In Fig. 3a–d the distribution of the largest Lyapunov exponent is given for four combinations of number of neurons, n , and dimension, d , with the spread held constant at $s = 8$. Approximately 600 000 networks were used to construct these distributions. In Fig. 3c and d ($d = 64$) it can be seen that the definition of chaos of $\lambda > 0$ is not sensitive to the threshold value of $\lambda = 0$. The figures also show that as the dimension increases and as the number of neurons increases, the distribution function of the largest Lyapunov exponent shifts to the right, eventually ending up entirely, but barely, to the right of the origin.

A prominent feature, most evident at low dimension, is the sharp spike at $\lambda = 0$, corresponding to quasi periodic orbits. The area under this spike is relatively small, suggesting that such orbits are a small fraction of the total. The fraction is less than the 30–40% found by Sprott (1993) for polynomial maps, but similar to the 5–10% found for low-dimensional polynomial flows.

5. Conclusion

The fact that our artificial neural networks tend to be chaotic with a probability approaching unity as their complexity increases is interesting, but perhaps

⁶ See Doyon et al. (1993) for a discussion of the route to chaos as the scaling parameter, s , increases.

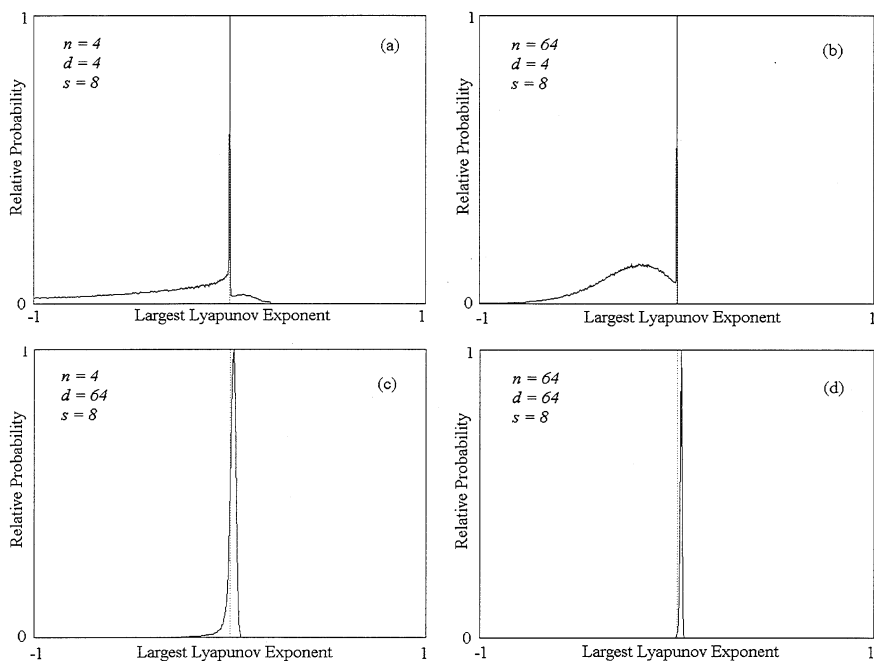


Fig. 3. Distributions of largest Lyapunov exponents for various n and d with $s = 8$.

not surprising. Less obvious is the observation that the largest Lyapunov exponent, though positive, approaches zero in the limit of increasing complexity. That is to say, our networks tend to be chaotic but only weakly so. They lie at the transition between ordered and chaotic behavior, sometimes called ‘the edge of chaos’. It has been conjectured that complex emergent systems (biological and otherwise) evolve toward such a phase transition (Langton et al., 1992).

Our results can be interpreted to suggest that evolution is not essential, but that such behavior is inherent in the system itself as its complexity increases. Our networks automatically exhibit both unpredictability (sensitivity to initial conditions) and long-term memory (small Lyapunov exponents), which are characteristics of complex systems not only in economics but in many other fields. However, it may also be that as the dimension increases, the networks average over more past values of the variable, and thus the time scale for the dynamics is increased, resulting in smaller values of the Lyapunov exponent.⁷ Each point in the time series is a weighted average of the previous d points. Thus the time scale

⁷ For more detailed discussion of the effect of dimension see Albers et al. (1996).

for the dynamics ought to become longer as d increases. The change in various dynamical quantities per iteration, including the rate of separation of nearby orbits, should therefore decrease. This expectation was tested by calculating the sum of the Lyapunov exponents given by the time average of

$$\ln \left| \frac{df}{\partial y_1} \right|.$$

Although this sum was strongly negative, implying rapid contraction of the phase-space volume, the average Lyapunov exponent, obtained by dividing the sum by d , approached zero with increasing d in about the same way as did the largest Lyapunov exponent.

In terms of economic dynamics, the work of Boldrin and Montrucchio (1986) in the infinite horizon growth model and Carrera and Moran (1993) in the overlapping generations model shows that any dynamical system is the equilibrium of some economy.⁸ Large economies characterized by non-linear dynamics are thus likely to have chaotic dynamics.

Acknowledgements

The authors would like to thank Derek Wright of the U.W. Madison computer science Condor group for help collecting data. Much of the data was run via the Condor High-Throughput Computing System; without this resource our data collection would be far more sparse. Part of this work was supported by the University of Wisconsin Hilldale Foundation.

References

- Albers, D.J., Sprott, J.C., Dechert, W.D., 1996. Dynamical behavior of artificial neural networks with random weights. In: Dagli, C.H., Akay, M., Chen, C.L.P., Fernandez, B.R., Ghosh, J., Intelligent Engineering Systems Through Artificial Neural Networks, Artificial Neural Networks in Engineering, vol. 6, ASME Press, New York, pp. 17–22.
- Boldrin, M., Montrucchio, L., 1986. On the indeterminacy of capital accumulation paths. *Journal of Economic Theory* 40, 26–39.
- Brock, W.A., 1993. Pathways to randomness in the economy: emergent nonlinearity and chaos in economics and finance. *Estudios Economicos* 8, 3–55.
- Carrera, C., Morán, M., 1993. General dynamics in overlapping generations models. Instituto Complutense de Análisis Económico, Universidad Complutense, Madrid.
- Doyon, B., Cessac, B., Quoy, M., Samuelides, M., 1993. Control of the transition to chaos in neural networks with random connectivity. *International Journal of Bifurcation and Chaos* 3, 279–291.

⁸ The known cases involve an unrealistically low discount factor. However, work by Sorger (1992) shows that the size of the discount factor can be considerably larger than the ones in the known examples.

- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366.
- Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks* 3, 535–549.
- Langton, C.G., Taylor, C., Farmer, J.D., Rassmussen, S. (Eds.), 1992. *Artificial Life II*, Santa Fe Institute Studies in the Sciences of Complexity, vol. 10, Addison-Wesley, Redwood City, CA.
- l'Écuyer, P., 1988. Efficient and portable combined random number generators. *Communications of the ACM* 31, 742–749.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992. *Numerical Recipes in C*, 2nd ed. Cambridge University Press, Cambridge.
- Sorger, G., 1992. Minimum Impatience Theorems for Recursive Economic Models, *Lecture Notes in Economics and Mathematical Systems*, vol. 390. Springer, Berlin.
- Sprott, J.C., 1993. How common is chaos? *Physics Letters A* 173, 21–24.
- Takens, F., 1981. Detecting strange attractors in turbulence. In: Rand, D., Young, L. (Eds.), *Dynamical Systems and Turbulence*, Warwick, 1980, *Lecture notes in Mathematics*, vol. 898. Springer, Berlin, pp. 366–381.