Distinguishing Nonlinear Stochasticity from Linear Stochasticity and Nonlinear Determinism

Yoshito Hirata
Mathematics and Informatics Center, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Graduate School of Information Science and Technology, The University of Tokyo,
Bunkyo-ku, Tokyo, Japan
International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo,
Bunkyo-ku, Tokyo, Japan
E-mail: hirata@mist.i.u-tokyo.ac.jp

Masanori Shiro
Mathematical Neuroinformatics Group, Advanced Industrial Science and Technology,
Tsukuba, Ibaraki 305-8568, Japan

Abstract

We propose two tests for distinguishing nonlinear stochasticity from linear stochasticity and nonlinear determinism based on a time series. One method makes a distinction between stochasticity and determinism using permutations, while the other method highlights the differences between linearity and nonlinearity using surrogate data and test statistics, which are not directly related to determinism or predictability. We analyze the validity of these two tests using toy models.

1 Introduction

In general, systems can be divided into four classes based on two axes (Fig. 1): whether they are deterministic or stochastic, and whether they are linear or nonlinear. Although linear deterministic systems, linear stochastic systems, and nonlinear deterministic systems have been thoroughly investigated, there is a limited number of time series analysis methods that can characterize nonlinear stochasticity. This difficulty arises from the history of nonlinear time series analysis, where studies have compared nonlinear determinism to linear stochasticity using surrogate data [1], which assumes linear stochasticity, based on test statistics characterizing nonlinear determinism, such as the parallelism of neighboring orbits [2] and predictability [1].

When we define underlying system dynamics as a map, the dynamics are called deterministic if the map does not include any probabilistic components. Otherwise, the dynamics are called stochastic. Additionally, the dynamics are called linear if the map is a linear function in terms of system variables. Otherwise, the underlying dynamics are called nonlinear.

If we are interested in analyzing deterministic-stochastic tendencies, there are several methods in the existing literature. For example, the “determinism vs. stochasticity” algorithm [3] utilizes a local linear model and varies the size of neighborhoods. If small neighborhoods yield good prediction, then the underlying dynamics are assumed to be deterministic and nonlinear. If prediction error is minimized by large neighborhoods, then the underlying dynamics are high dimensional and can be either stochastic or deterministic. The methods proposed by Kaplan and Glass [4], and Wayland et al. [2] focus on the parallelism of neighboring orbits. If neighboring orbits are parallel, then the underlying dynamics are assumed to be deterministic. Otherwise, stochasticity is inferred. In delayed vector variance [5], the sum of normalized prediction errors is evaluated based on varying the size of the span for neighborhoods. Rosso et al. [6] proposed the complexity-entropy-causality plane to distinguish correlated noise from deterministic systems. If a system is stochastic, then the time series is located at the bottom-right corner of the complexity-entropy-causality plane [6]. Amigó et al. [7] used the property that the number of
appearing ordinal patterns grows exponentially when the length of ordinal patterns is enlarged for a deterministic system \[8\]. This tendency indicates that there are missing ordinal patterns in a time series generated by a deterministic system. Such missing ordinal patterns can be used to infer the randomness of the underlying dynamics \[7\]. In horizontal visibility graphs \[9\], a time series is converted into a network. Various network observations are made, such as the degree distribution and average degree of nodes, to construct a test to distinguish uncorrelated random series from chaotic time series \[9\]. However, these methods are either qualitative or quantitative based on uncorrelated random series or linear surrogates. Therefore, to the best of our knowledge, there are no methods that quantitatively test stochasticity without relying on linear properties.

Therefore, we separated the axis of linearity-nonlinearity from that of determinism-stochasticity and developed a set of methods for testing the properties of each axis independently. The key idea of this approach is to define a test statistic that is independent of determinism for characterizing the nonlinearity of a system.

### 2 The proposed set of methods

Here, we describe each of the proposed tests individually.

#### 2.1 Deterministic or stochastic?

We use permutations \[10, 7\] to determine if a given time series is generated from a deterministic or stochastic system. A permutation is a rank order of time points for a segment of a time series. For example, if a segment of a time series is \[2.5, 3, -0.5, 1\], then its rank-order is \[3, 4, 1, 2\] because we can sort it in ascending order as \[-0.5, 1, 2.5, 3\]. Then, we treat \((3, 4, 1, 2)\) as a pattern for a symbol subsequence.

When we know the length \(L\) of the permutations, we can count the number \(N(L)\) of different permutations in a given time series. Amigó et al. \[7\] argued that the number \(N(L)\) increases exponentially when \(L\) increases if the underlying dynamics are deterministic. If the underlying dynamics are stochastic, then \(N(L)\) increases in a factorial or combinatorial fashion. Therefore, we identify these differences in \(\log N(L)\) using the \(F\)-distribution \[11\]. If the underlying dynamics are deterministic and we use \(L = 3, \ldots, L_{\text{max}}\), then there are \(L_{\text{max}} - 4\) degrees of freedom. However, if the underlying dynamics are stochastic, there are \(L_{\text{max}} - 2\) degrees of freedom to determine if the underlying dynamics are deterministic or stochastic.

Some small examples are presented below. The first example is the following logistic map \[12\]:

\[
x(t + 1) = 3.8x(t)(1 - x(t)),
\]

where we generated 20 time series of length 100000.

The results for this map are presented in Fig. 2. The error bars obtained from 20 simulations cover the linear fitting between lengths of four and seven (the length of three was excluded from the fitting because Fig. 2 indicates that the length of three is too short to maintain exponential growth). Therefore, we cannot reject the null hypothesis that the underlying dynamics are deterministic in nature.

The second example is the following autoregressive linear (AR) model \[13\]:

\[
x(t + 1) = -0.8x(t) + \eta(t),
\]

where \(\eta(t)\) follows a Gaussian distribution with a mean of 0 and standard deviation of 1.

The results for this model are presented in Fig. 3. Although it is difficult to see, the error bars in this case are very narrow and exclude linear fitting results outside of the 95% confidence intervals. Therefore, in this case, we can reject the null hypothesis that the underlying dynamics are deterministic in nature. It is
more natural to consider a time series generated from a stochastic system.

Our third example is that of the GARCH model [14]. We consider a model identical to that described in [14]. The results presented in Fig. 4 reveal patterns similar to those in Fig. 3. Specifically, the null hypothesis that the number of ordinal patterns increases exponentially was rejected. Therefore, the first proposed test behaves properly.

If we use the above method directly, there is a problem when the underlying dynamics exhibit period-doubling bifurcation or an approximately super-stable periodic cycle. Therefore, we subsample a given time series every 1, 2, 3, 4 time points and apply the method above indirectly. The significance level threshold must be adjusted appropriately.

2.2 Linear or nonlinear?

We utilize truncated Fourier transform surrogates [15] to determine if underlying dynamics are linear or nonlinear. A common practice is to choose a test statistic that can characterize nonlinear determinism, such as the parallelism of neighboring orbits and prediction errors. However, to distinguish the test for nonlinearity from the test for stochasticity completely, we use a mean of \( x(t)^2 x(t+c)^2 \) as a test statistic for \( c = 1, 2, \ldots, 100 \). Because we use Fourier-transform-based surrogates that assume the stationarity of a time series [16], we also assume that a given time series is stationary.

We generate 39 surrogates such that each test represents a significance level of 5%. We then assume the independence of each test and apply a binomial distribution of the number 100 and a probability of 0.05 to calculate the overall \( p \)-value.

3 Numerical examples

We generated time series using the following model:

\[
x(t+1) = a(-2.7x(t) + 3.7x(t)^3) + (1-a)(-0.8x(t) + \sigma u(t)),
\]

where \( \sigma = 0.1 \) and \( u(t) \sim N(0, 1) \) for each \( t \). When \( a = 0 \), the model coincides with an autoregressive linear model [13], which is a typical model for linear stochasticity. When \( a = 1 \), the model becomes a nonlinear deterministic map called a cubic map [17]. Between these values, stochasticity and nonlinearity may coexist.

The bifurcation diagram and scatter plots corresponding to this model are presented in Figs. 5 and 6, respectively. One can see period-doubling bifurcation at approximately \( a = 0.1 \) and a periodic solution with a period of three at \( a = 1 \).

For each \( a \), we generated a time series of length 100000.

The results are presented in Fig. 7. When \( a \leq 0.99 \), one can see stochasticity. Additionally, one can see non-
Fig. 7: Results of tests for (a) stochasticity and (b) nonlinearity for the model in Eq. (3). In each panel, the red solid line was obtained from the proposed tests and the blue dashed line corresponds to a significance level of 5%.

linearity when $a \geq 0.04$. Therefore, at $0.04 \leq a \leq 0.99$, one could detect nonlinear stochasticity. These results make sense due to the construction of this model.

Another example is the following model:

$$x(t + 1) = a(1 - 1.8x(t)^2) + (1 - a)(-0.8x(t) + \sigma u(t)). \tag{4}$$

This model connects an AR model [13] and logistic map [12] using a parameter $a$. The other settings are similar to the previous case.

We present a bifurcation diagram for the model depending on the parameter $a$ in Fig. 8. Additionally, we present scatter plots for a set of the parameters in Fig. 9. One can see that before $a$ reaches 0.60, the equilibrium point is unstable and a period of two appears. Next, before $a$ reaches 1, a periodic solution with a period of three appears.

The results are presented in Fig. 10. When $a \geq 0.41$, we present a bifurcation diagram for the model in Eq. (4). See the caption of Fig. 7 to interpret the results.

Fig. 8: Bifurcation diagram for the model in Eq. (4).

Fig. 9: Scatter plots for the model in Eq. (4) based on the parameter $a$.

Fig. 10: Results of tests for (a) stochasticity and (b) nonlinearity for the model in Eq. (4).
nonlinearity is detected, except for the value of $a = 0.94$. The test for stochasticity is satisfied when $a \leq 0.85$, $a = 0.97$, and $a = 0.98$. Therefore, for $0.41 \leq a \leq 0.85$, nonlinear stochasticity is detected. When the mixing parameter $a$ is in the middle range, nonlinear stochasticity is evident. These results agree with our expectation for constructing this model.

4 Discussion

Here, we first analyze the robustness of the proposed tests against time series length and observation noise. We evaluated robustness using three models: the logistic map, AR model, and GARCH model.

The results are presented in Fig. 11. Regarding the test for stochasticity, the length of a time series should be at least 100000. When the time series was shorter, the proposed method did not work well in the case of the AR model (Fig. 11c). Additionally, observational noise should be 5% or less (Fig. 11).

The test for linearity seems to work well if the length of a time series is 5000 or longer (Fig. 11b,d,f). The test for linearity is robust against observational noise in the tested range. Overall, the proposed tests require a time series length of 100000 with low noise.

We will now analyze the applicability of the proposed tests to high-dimensional systems. We applied the proposed tests to a time series of length 100000 that was generated from a 240-dimensional Lorenz’96 II model [18,19]. We found that this dataset is likely to be deterministic and nonlinear. These results mean that the proposed tests did not consider the 240-dimensional Lorenz’96 II model to be stochastic. This result may be a promising indicator for the proposed tests and should be investigated more deeply in the future.

We noticed that the proposed framework does not work well when we apply the proposed tests to datasets generated from the model of noise-induced order [20]. For example, the results in Fig. 12 look more similar to the results in Fig. 2 than those in Figs. 3 and 4. However, it should be noted that the values in Fig. 12 are much smaller than those in Figs. 2-4. In such cases, although nonlinearity is detected, stochasticity is not detected because of the regularity generated by noise. The application of pseudo-periodic surrogates [21] with Wayland statistics [2] reveals that in this model, there is determinism beyond any apparent periodicity (see Fig. 13, p-value: 0.0022).

In general, noise can lead to either of the following two phenomena: (i) noise increases complexity or (ii) noise decreases complexity. In case (i), the proposed tests seem to work well. However, in case (ii), the proposed tests do not provide appropriate results. We must prepare another test or set of tests for handling case (ii). We are currently working in this direction and will likely present the results in a future communication.

Additionally, the model of noise-induced order is apparently correlated. Therefore, to overcome the prob-
lems discussed above, we must derive a method that can handle highly correlated time series as a null hypothesis. This point will be also covered in a future communication.

After identifying that underlying dynamics are nonlinear and stochastic, modeling such dynamics will be our next major focus.

It should be noted that if underlying dynamics contain one or more time-varying parameters, then the target system is non-stationary, leading to a spurious rejection of the null hypothesis of linearity [16]. Therefore, a test for stationarity should be applied before one analyzes a given time series using the proposed method.

In this paper, we proposed a set of tests for identifying the nonlinear stochasticity of a time series. A test for stochasticity was constructed based on permutations and a test for nonlinearity was constructed based on surrogate data and nonlinear test statistics. Numerical examples demonstrated that the proposed tests perform correctly by detecting nonlinear stochasticity in the middle range of a parameter connecting an autoregressive model to a nonlinear map. We believe that the proposed tests will expand a new frontier for research on nonlinear stochasticity.

Acknowledgement

The research by Y.H. is supported by JSPS KAKENHI Grant Number JP18K11461.

References

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer: Testing for nonlinearity in time series: the method of surrogate data, Physica D, 58, pp. 77-94, 1992.

[2] R. Wayland, D. Bromley, D. Pickett, and A. Pasamante: Recognizing determinism in a time series, *Phys. Rev. Lett.*, 70, pp. 580-582, 1993.

[3] M. C. Casdagli and A. S. Weigend: Exploring the continuum between deterministic and stochastic modeling, In Eds. A. S. Weigend and N. A. Gershenfeld, *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, pp. 347-366, 1993.

[4] D. T. Kaplan and L. Glass: Direct test for determinism in a time series, *Phys. Rev. Lett.*, 68, pp. 427-430, 1992.

[5] T. Gautama, D. P. Mandic, and M. M. Van Hulle: The delay vector variance method for detecting determinism and nonlinearity in time series, *Physica D*, 190, pp. 167-176, 2004.

[6] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes: Distinguishing noise from chaos, *Phys. Rev. Lett.*, 99, art. no. 154102, 2007.

[7] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán: Combinatorial detection of determinism in noisy time series, *EPL*, 83, art. no. 60005, 2009.

[8] J. M. Amigó and M. B. Kennel: Topological permutation entropy, *Physica D*, 231, pp. 137-142, 2007.

[9] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque: Horizontal visibility graphs: Exact results for random time series, *Phys. Rev. E*, 80, art. no. 046103, 2009.

[10] C. Bandt and B. Pompe: Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.*, 88, art. no. 174102, 2002.

[11] M. H. DeGroot and M. J. Schervish: *Probability and Statistics*, Addison Wesley, 2001.

[12] R. M. May: Simple mathematical models with very complicated dynamics, *Nature*, 261, pp. 459-467, 1976.

[13] J. D. Hamilton: *Time Series Analysis*, Princeton University Press, 1994.

[14] C. G. Lamoureux and W. D. Lastrapes: Persistence in variance, structural change, and the GARCH model, *J. Business & Economic Statistics*, 8, pp. 225-234, 1990.

[15] T. Nakamura, M. Small, and Y. Hirata: Testing for nonlinearity in irregular fluctuations with long-term trends, *Phys. Rev. E*, 74, 026205, 2006.

[16] J. Timmer: Power of surrogate data testing with respect to nonstationarity, *Phys. Rev. E*, 58, 5153-5156, 1998.

[17] B.-L. Hao: *Elementary Symbolic Dynamics and Chaos in Dissipative Systems*, World Scientific, 1989.

[18] E. N. Lorenz: Predictability: A problem partly solved. In *Proceedings of the Seminar on Predictability*, Vol. 1, pp. 1-18 ECMWF, 1996.

[19] J. A. Hansen and L. A. Smith: The role of operational constraints in selecting supplementary observations, *J. Atmos. Sci.*, 57, pp. 2859-2871, 2000.

[20] K. Matsumoto and I. Tsuda: Noise-induced order, *J. Stat. Phys.*, 31, pp. 87-106, 1983.

[21] M. Small, D. Yu, and R. Harrison: Surrogate test for pseudoperiodic time series data, *Phys. Rev. Lett.*, 87, art. no. 188101, 2001.