

Bootstrap accuracy estimation in chaotic time series ¹

Stima bootstrap dell'accuratezza per serie storiche caotiche

Simone Giannerini e Rodolfo Rosa

Dipartimento di Scienze Statistiche “Paolo Fortunati”

Università di Bologna

giannerini@stat.unibo.it

Riassunto: Proponiamo un nuovo metodo basato sul ricampionamento per stimare l'accuratezza delle stime per serie storiche caotiche continue. Si dimostra con esperimenti al calcolatore che lo schema suggerito permette di generare serie che preservano la dipendenza lineare e non lineare dei dati originali. Inoltre, la distribuzione del massimo esponente di Lyapunov globale, ottenuta con il nostro metodo a partire da una singola serie, è in accordo con quella che si ottiene da serie “vere” generate con differenti condizioni iniziali. Gli esperimenti sono eseguiti sia con serie “pulite” che contaminate da rumore di osservazione.

Keywords: Chaotic time series, bootstrap, invariants estimation.

1. Introduction

Chaos theory offers new perspectives for time series analysis as well as concepts and ideas that carry remarkable contributions to statistics. At the same time, statistical methodology has shown to play a crucial role for the comprehension of nonlinear and chaotic phenomena.

We presented two years ago at SIS 2000 a resampling method based on spline interpolation to face the problem of pointwise and interval estimation in time-continuous chaotic time series (Giannerini and Rosa, 2000). There we have shown that the observations obtained by resampling do not differ significantly from the true observations obtained by integrating the differential equations of the system at the same time points. Here we present an improved scheme that allows the derivation of meaningful confidence intervals for invariant quantities. In particular, we will show that the distribution of the global maximal Lyapunov exponent (MLCE) obtained by way of resampling a single series with our method, agrees with the true distribution obtained from series with different initial conditions. The method works both with clean series and in presence of measurement noise.

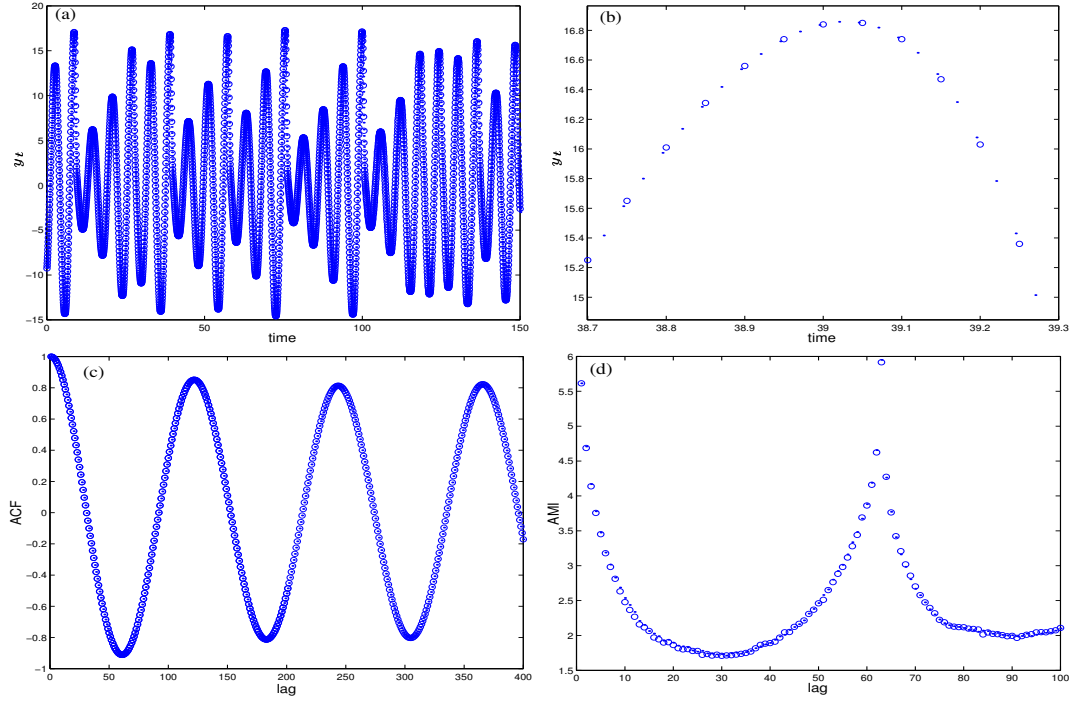
2. Method

Suppose we have a continuous time dynamical system in $\mathbb{R}^d$ described through a set of d ordinary differential equations:

$$\frac{d}{dt}\mathbf{x}(t) = F(\mathbf{x}(t)), \quad t \in \mathbb{R}. \quad (1)$$

¹Funding from MURST is gratefully acknowledged

Figure 1: True (Open circles) and resampled (dots): (a) time series; (b) enhanced particular of (a); (c) autocorrelation function; (d) average mutual information.



where $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\mathbf{x} \in \mathbb{R}^d$. The states of the system are not observed directly but through an observation function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, a (nonlinear) projection that defines the scalar time series we observe in practice: $y_t = f(\mathbf{x}(t)) + s\epsilon_t$. Here $\{\epsilon_t\}$ is a white noise process supposed to represent measurement error and s is a constant that determines the signal to noise ratio of variances (SNR). Thus, we are given a univariate experimental time series $\mathbf{y} = \{y_{1,\Delta t}, \dots, y_{n,\Delta t}\}$ of length n and sampling rate Δt . From this series the state space dynamics of the system is reconstructed by means of an appropriate choice of the embedding dimension m and time delay τ . The resampling procedure may be described as follows:

1. A new sampling rate δt is derived by dividing Δt by a factor k .
2. A point t_1 in the interval $]\Delta t, 2\Delta t[$ is selected at random.
3. A new series $\mathbf{y}_1^r = y_1^r, y_2^r, \dots, y_{q+1}^r$, (r stands for “replicated”) is obtained by interpolating the original series at the points $(t_1, t_1 + 1 \cdot \delta t, \dots, t_1 + q \cdot \delta t)$ where q is the last point less than $n \times k$, that is:

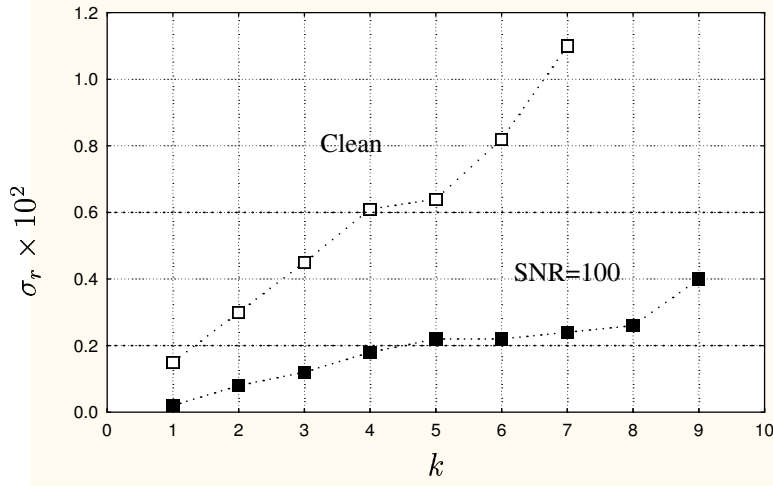
$$q = n \cdot k - \left[t_1 \cdot \frac{k}{\Delta t} \right]$$

where $[\cdot]$ indicates the integer part of a number.

4. Repeat steps 2 and 3 B times, obtaining B resampled series $\mathbf{y}_1^r, \mathbf{y}_2^r, \dots, \mathbf{y}_B^r$. On each of these series it is possible to estimate a statistic of interest, obtaining its empirical distribution.

In Giannerini and Rosa (2001) we have performed extensive computer experiments to assess the validity of the method. Here focus will be given to two main points: (i)

Figure 2: Behaviour of σ_r vs. k for the Rössler attractor. Clean series (open squares) and with SNR=100 (full squares).



agreement of true and resampled series with respect to linear and non linear dependence; (ii) agreement of the true and bootstrap distribution of the MLCE. We chose to perform experiments on the MLCE since it is an important measure of stability and one of the indicators of the presence of chaos, even though the method may be applied to other statistics of interest as well.

3. Results and discussion

As an example we report results from the x component of the Rössler attractor at chaotic regime:

$$dx/dt = -(y + z); \quad dy/dt = x + ay; \quad dz/dt = b + z(x - c); \quad (2)$$

where $a = 0.15$, $b = 0.20$, $c = 10.0$, $\Delta t = 0.1$, $m = 3$, $\tau = 8$, length of the series $n = 6500$ and number of replications $B = 100$. In fig. 1 (a) we show the true series together with a resampled one. Clearly, the two plots are indistinguishable. The enhanced particular of a peak is displayed in fig. 1 (b). It appears that our resampling procedure does not cause a reduction in the dynamical range of the signal, due to the use of splines, as also remarked in Kantz and Schreiber (1997). Both linear and nonlinear dependence are preserved by the resampling mechanism as it is displayed in fig. 1 (c) autocorrelation function, 1 (d) average mutual information. Also in these instances the plots appear almost superimposed. The same agreement is found from the inspection of the periodogram and the power spectral density (not shown here).

Figure 2 shows the behaviour of the bootstrap standard deviation of the MLCE's distribution σ_r as a function of the factor k for both clean and noisy series (SNR = 100). For an analysis regarding several values of SNR (including SNR = 50) see Giannerini and Rosa (2001). Notice that for SNR < 100 the stochastic component tends to overwhelm the deterministic structure making estimation unreliable (see also Kantz and Schreiber 1997, pag. 78). From the figure it results that the σ_r undergoes an increase, stabilizes on a plateau and then displays again a steep rise. On the plateau σ_r coincides with σ_{tr} the true standard deviation of the MLCE's distribution (dashed horizontal lines). This finding has been confirmed by means of parametric and nonparametric statistical tests in which

we compare true and bootstrap distributions of the MLCE, as shown in Tab. 1 and Tab. 2. We have chosen (i) the Kolmogorov-Smirnov two sample test to assess the agreement of the distributions (ii) the Levene test and the Brown-Forsythe test for the homogeneity of variance and (iii) the standard F test since MLCE's distribution turned out to be normal in most cases.

Table 1: *Clean series; p-values for the following tests: Kolmogorov-Smirnov (K-S), F, Levene, Brown and Forsythe (B-F).*

k	2	3	4	5	6	7
K-S	$p < .03$	$p > .10$	$p > .10$	$p > .10$	$p > .10$	$p < .02$
F	0.00	0.00	0.45	0.51	0.00	0.00
Levene	0.00	0.00	0.21	0.22	0.01	0.00
B-F	0.00	0.00	0.28	0.29	0.02	0.00

In Tab. 1 (clean series) the p -values pertaining to the above mentioned tests are given. Clearly, for $k = 4$ and $k = 5$ all the performed tests do not reject the null hypothesis of equality of the variances/distributions. The same behaviour, found in Tab. 2 (SNR=100) for k ranging from 4 to 7, confirms the visual impression.

Table 2: *As in Table 1 but for noisy series (SNR=100).*

k	2	3	4	5	6	7	8	9
K-S	$p < .01$	$p > .10$	$p > .10$	$p > .10$	$p > .10$	$p > .10$	$p > .10$	$p < .03$
F	0.00	0.00	0.34	0.88	0.61	0.14	0.04	0.00
Levene	0.00	0.00	0.29	0.84	0.95	0.30	0.03	0.00
B-F	0.00	0.00	0.28	0.85	0.95	0.42	0.04	0.00

We may conclude that our scheme, being independent from the procedure of reconstruction of the dynamics in $\mathbb{R}^m$, does not suffer from the presence of influential points and from recursive exponentiation of errors which plague methods based upon estimation in phase space. Moreover the “fattening” of the series by a factor k allows one to exploit all the information at disposal, without running into sources of bias.

References

- Giannerini S. and Rosa R. (2000) Is resampling of chaotic time series feasible?, in: *XL conference of the Italian Statistical Society*, Firenze, 26-28 Aprile, 295–298.
- Giannerini S. and Rosa R. (2001) New resampling method to asses the accuracy of the maximal lyapunov exponent estimation, *Physica D*, 155, 101–111.
- Kantz H. and Schreiber T. (1997) *Non linear time series analysis*, Cambridge University Press, Cambridge.