Some recent nonparametric tools for time series data analysis

Howell Tong

University of Hong Kong, Hong Kong, China and University of Kent, UK.

Introduction and Summary

We have recently developed a number of non-parametric techniques to aid our analysis of physical data. Many of these techniques parallel developments in deterministic dynamical systems including chaos. For reasons of space, we shall outline here some of those techniques which we have found useful in practical applications but omit their detailed theoretical justifications, which may be obtained in the references cited and the forthcoming book by Chan and Tong (1999). This paper is a continuation of developments reported in Tong (1995). Numerical illustrations will be given in the verbal presentation.

Background

We use as our basic setup a nonlinear autoregressive (NLAR) model, namely

\[ Y_t = f_d(Y_{t-1}, \ldots, Y_{t-d}) + \epsilon_t, \]

where \( f_d \) is typically unknown and \( \{ \epsilon_t \} \) is a sequence of iid random variables with zero mean and finite variance. Note that the assumption of independence may be weakened. In a deterministic dynamical system (or DDS for short), \( var(\epsilon_t) = 0 \). Given data \( \{ Y_1, Y_2, \ldots, Y_n \} \), one of the first tasks is to determine \( d \), which is called the order of the NLAR model and is generally unknown. It corresponds to the so-called embedding dimension in a DDS. One of the more successful techniques there is the method of false nearest neighbours. Our approach is to use a cross-validatory approach consisting of two basic steps: (i) using a Nadaraya-Watson (N-W) estimate (essentially a locally constant estimate) of \( f_d \) based on the non-deleted data to forecast the deleted datum and (ii) repeating the procedure across the whole sample by deleting one datum at a time. We finally minimise the sum of squares of the prediction errors, called the cross-validatory sum of squares (CVRSS(d)) with respect to the tuning parameter (called the bandwidth) for the N-W estimate (conditional on \( d \)) and then \( d \), to obtain the cross-validatory estimate of the order, which, under mild conditions, is consistent. Details are given in Cheng and Tong (1992). In what follows, we shall assume that \( d \) is given, denote \( (Y_{t-1}, \ldots, Y_{t-d})^T \) by \( X_t \) and write \( f_d \) simply as \( f \) unless there is danger of confusion.

Detection of operational determinism

Yao and Tong (1998) have developed a parallel of the deterministic versus stochastic test developed by Casdagli (1992). The objective is to detect if \( var(\epsilon_t) \) is close to zero. Recall that the locally constant estimate of \( f \) may be extended to the locally linear estimate by minimising w.r.t. \( a \) and \( b \)

\[ \sum_{t=1}^{n} \left| Y_t - a - b^T(X_t - x) \right|^2 K \left( \frac{X_t - x}{h} \right), \]
where $K(\cdot)$ is a kernel function and $h > 0$ is the bandwidth. The minimising $a$ and $b$ give respectively the locally linear estimate, say $\hat{f}_{n,h}$, of $f$ and its derivative. Now, there are several ways to estimate $h$, the cross-validation method being one of the common ones. Yao and Tong (op. cit.) have suggested an alternative method which is quite fast: split the sample into two subsets, use the first subset of $m$ observations to estimate $f$ and then minimize the weighted sum of squares of predictions on the second subset of $n - m$ observations, namely

$$\sum_{t=m+1}^{n} \{Y_t - \hat{f}_{m,h}(X_t)\}^2 W(X_t),$$

where $W(\cdot)$ is a weight function. Let $\hat{h}_m$ denote the minimiser. We set the estimate for the bandwidth for the whole sample as

$$\hat{h}_n = \left(\frac{m}{n}\right)^{1/(d+4)} \hat{h}_m.$$

The Yao-Tong detection goes as follows.

1. For the given data $\{(X_t, Y_t), 1 \leq t \leq n\}$, obtain the estimate $\hat{h}_n$ as given above.
2. Obtain the locally linear regression estimator $\hat{f}_{n,h}(\cdot)$ using $h = \hat{h}_n$, and calculate the residuals $\hat{\epsilon}_t = Y_t - \hat{f}_{n,h_n}(X_t)$ for $t = 1, \ldots, n$.
3. Bootstrap: draw $n$ independent random numbers $i_1, \ldots, i_n$ from the uniform distribution with the sample space $\{1, \ldots, n\}$, and define $\epsilon^*_t = \hat{\epsilon}_{i_t}$ for $t = 1, \ldots, n$. Form the bootstrap sample $\{(X_t, Y^*_t), 1 \leq t \leq n\}$ with

$$Y^*_t = \hat{f}_{n,h_n}(X_t) + \epsilon^*_t.$$

4. Obtain an estimate $\hat{h}^*_n$ from the sample $\{(X_t, Y^*_t), 1 \leq t \leq n\}$ as in Step 1. Especially, the search for $\hat{h}^*_n$ around $\hat{h}_n$ is conducted on finer grids than those used in Step 1.
5. Repeat Steps 3 and 4 $N$ times, and count the frequency of occurrence of the event that $\hat{h}^*_n \leq \hat{h}_n$. Then the relative frequency $\alpha (= \text{frequency}/N)$ is taken as a measure of how plausible it is that the data are generated by an operationally deterministic model.

**Initial value sensitivity**

Dynamicists typically use the so-called *Lyapunov exponent* to measure initial-value sensitivity, which is central to the understanding of chaos. In our experience, we have found $b$ described in the last section a practically useful parallel to the Lyapunov exponent, in the sense that large values of $b$ tend to indicate those parts of the time series (more precisely state space) over which the conditional mean function (i.e. the least-square predictor) is sensitive to initial values. For details, see Yao and Tong (1994a, 1994b)
Estimating the conditional mean function for high dimensional data

It is well known that the estimation of $f$ is plagued by the curse of dimensionality, namely the size of $d$. In practice, a linear model is often a good first approximation. Recently, Xia, Tong and Li (1997) have proposed a generalized partially linear single-index model which combines the advantage of a parametric linear model and the flexibility of a nonparametric nonlinear model but at the same time moderate the impact of the curse. In the time series context, their model amounts to expressing

$$f(X_t) = \beta^T X_t + \phi(\theta^T X_t)$$

where $\beta$ and $\theta$ are unknown parametric vectors perpendicular to each other, and $\phi(\cdot)$ is an unknown function. Under mild conditions, they have obtained estimates of $\beta$ and $\theta$ which are root-$n$ consistent. They have also obtained an estimate of $\phi(\cdot)$ which has a uniform convergence rate of $(n^{4/5} / \ln n)^{1/2}$.

Concluding Remarks

Nonparametric methods have provided the statisticians with powerful tools to tackle many of the problems of vital importance to the dynamical system approach to physical data. We have sketched some of these tools. Space has not permitted us to describe our other tools. These include a test for multi-modality for time series data (Chan and Tong, 1998), which should aid the identification for multiple equilibria in a DDS; estimation for an instantaneous nonparametric transformation for time series (Xia, Tong, Li and Zhu, 1998), which should aid the search for an appropriate re-scaling in a DDS; a test for symmetries of multivariate probability distributions (Diks and Tong, 1999), which is relevant to the understanding of chaos with symmetry.

References


