ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bootstrap prediction intervals for Markov processes

Li Pan, Dimitris N. Politis *,[1]

Department of Mathematics, University of California-San Diego, La Jolla, CA 92093-0112, USA

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 30 May 2015
Accepted 30 May 2015
Available online xxxx

Keywords:
Confidence intervals
Local Bootstrap
Model-Free Prediction

ABSTRACT

Given time series data $X_1, \ldots, X_n$, the problem of optimal prediction of X_{n+1} has been well-studied. The same is not true, however, as regards the problem of constructing a prediction interval with prespecified coverage probability for X_{n+1}, i.e., turning the point predictor into an interval predictor. In the past, prediction intervals have mainly been constructed for time series that obey an autoregressive model that is linear, nonlinear or nonparametric. In the paper at hand, the scope is expanded by assuming only that $\{X_t\}$ is a Markov process of order $p \geq 1$ without insisting that any specific autoregressive equation is satisfied. Several different approaches and methods are considered, namely both Forward and Backward approaches to prediction intervals as combined with three resampling methods: the bootstrap based on estimated transition densities, the Local Bootstrap for Markov processes, and the novel Model-Free bootstrap. In simulations, prediction intervals obtained from different methods are compared in terms of their coverage level and length of interval.

1. Introduction

Prediction is a key objective in time series analysis. The theory of optimal – linear and nonlinear – point predictors has been well developed. The same is not true, however, as regards the problem of constructing a prediction interval with prespecified coverage probability, i.e., turning the point predictor into an interval predictor. Even in the related problem of regression, the available literature on prediction intervals is not large; see e.g. Geisser (1993), Carroll and Ruppert (1991), Olive (2007, 2015), Patel (1989), Schmoyer (1992), and Stine (1985). Recently, Politis (2013) has re-cast the prediction problem – including prediction intervals – in a Model-Free setting.

An autoregressive (AR) time series model, be it linear, nonlinear, or nonparametric, bears a formal resemblance to the analogous regression model. Indeed, AR models can typically be successfully fitted by the same methods used to estimate a regression, e.g., ordinary Least Square (LS) regression methods for parametric models, and scatterplot smoothing for nonparametric ones. There are several papers for prediction intervals for AR models (typically linear) that represent a broad spectrum of methods; see e.g. Alonso et al. (2002), Box and Jenkins (1976), Breidt et al. (1995), Masarotto (1990), Pascual et al. (2004), Thombs and Schucany (1990), and Wolf and Wunderli (2015).

Recently, Pan and Politis (in press) presented a unified approach towards prediction intervals when a time series $\{X_t\}$ obeys an autoregressive model that is either linear, nonlinear or nonparametric. We expand the scope by assuming only that $\{X_t\}$ is a Markov process of order $p \geq 1$ without insisting that any specific autoregressive equation is satisfied. Recall that Pan and Politis (in press) identified two different general approaches towards building bootstrap prediction intervals

* Corresponding author. Tel.: +1 858 534 5861; fax: +1 858 534 5273.
E-mail addresses: lipan@ucsd.edu (L. Pan), dpolitis@ucsd.edu (D.N. Politis).

[1] The website http://www.math.ucsd.edu/~politis/DPsoftware.html contains relevant software for the implementation of methods developed in this paper.

http://dx.doi.org/10.1016/j.csda.2015.05.010
0167-9473/

with *conditional validity*, namely the Forward and Backward recursive schemes. We will address both Forward and Backward approaches in the setting of Markovian data; see Section 2 for details.

In terms of the actual resampling mechanism, we will consider the following three options:

1. The bootstrap method based on kernel estimates of the transition density of the Markov processes as proposed by Rajarshi (1990); see Section 3.
2. The Local Bootstrap for Markov processes as proposed by Paparoditis and Politis (1998; 2002); see Section 4.
3. The Model-Free Bootstrap for Markov Processes; this is a novel resampling scheme that stems from the Model-Free Prediction Principle of Politis (2013). To elaborate, the key idea is to transform a given complex dataset into one that is i.i.d. (independent, identically distributed); having done that, the prediction problem is greatly simplified, and that includes the construction of prediction intervals. In the case of a Markov Process, this simplification can be accomplished using the Rosenblatt (1952) transformation; see Section 6.

In the case of time series that satisfy an autoregressive equation that is nonlinear and/or nonparametric, Pan and Politis (in press) noted that the Backward approach was not generally feasible. Recall that, under causality, AR models are special cases of Markov processes. Hence, in Section 5 we propose a *hybrid* approach for nonparametric autoregressions in which the forward step uses the autoregressive equation explicitly while the backward step uses one of the three aforementioned Markov bootstrap procedures.

In the following, Section 2 will describe the setting of the prediction problem under consideration, and the construction of bootstrap prediction intervals. All prediction intervals studied in the paper at hand are asymptotically valid under appropriate conditions. We will assess and compare the finite-sample performance of all the methods proposed via Monte Carlo simulations presented in Section 7. Appendix A is devoted to showing that a Markov process remains Markov after a time-reversal; this is needed to justify the use of all Backward bootstrap approaches. Finally, Appendix B discusses the problem of prediction intervals in r-step ahead prediction for $r \geq 1$.

2. Prediction and bootstrap for Markov processes

2.1. Notation and definitions

Here, and throughout the rest of the paper, we assume that $X = \{X_t, t = 1, 2, \ldots\}$ is a real-valued, strictly stationary process that is Markov of order p. Letting $Y_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})'$, we define

$$\begin{aligned} F(y) &= P[Y_p \leq y], \\ F(x, y) &= P[X_{p+1} \leq x, Y_p \leq y], \\ F(x|y) &= P[X_{p+1} \leq x | Y_p = y], \end{aligned} \tag{2.1}$$

for $x \in \mathbb{R},\ y \in \mathbb{R}^p$; in the above, we have used the short-hand $\{Y_p \leq y\}$ to denote the event {the ith coordinate of Y_p is less or equal to the ith coordinate of y for all $i = 1, \ldots, p$}.

Let $f(y), f(x, y), f(x|y)$ be the corresponding densities of the distributions in Eq. (2.1). We will assume throughout the paper that these densities are with respect to Lebesgue measure. However, our results in Sections 3 and 4, i.e., bootstrap based on estimated transition densities and Local Bootstrap, could be easily generalized to the case of densities taken with respect to counting measure; i.e., the case of discrete random variables. Remark 6.4 shows a modification that also renders the Model-Free bootstrap of Section 6 valid for discrete data.

Let $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ denote the observed sample path from the Markov chain X, and let $y_n = (x_n, \ldots, x_{n-p+1})'$. Denote by $\hat{X}_{n+1}$ the chosen point predictor of X_{n+1} based on the data at hand. Because of the Markov structure, this predictor will be a functional of $\hat{f}_n(\cdot|y_n)$ which is our data-based estimator of the conditional density $f(\cdot|y_n)$. For example, the L_2-optimal predictor would be given by the mean of $\hat{f}_n(\cdot|y_n)$; similarly, the L_1-optimal predictor would be given by the median of $\hat{f}_n(\cdot|y_n)$. To fix ideas in what follows will focus on the L_2-optimal predictor, usually approximated by $\hat{X}_{n+1} = \int x\hat{f}_n(x|y_n)dx$, with the understanding that other functionals of $\hat{f}_n(\cdot|y_n)$ can be accommodated equally well.

Remark 2.1. An integral such as $\int x\hat{f}_n(x|y_n)dx$ can be calculated by numerical integration, e.g. using the *adaptive quadrature* method. However, the L_2-optimal predictor can be approximated in several different ways that are asymptotically equivalent. The most straightforward alternative is a kernel smoothed estimator of the autoregression scatterplot, i.e., estimator (5.3) defined in the sequel. Remark 6.2 lists some further alternative options.

Beyond the point predictor $\hat{X}_{n+1}$, we want to construct a prediction interval that will contain X_{n+1} with probability $1 - \alpha$ asymptotically; the following definition is helpful.

Definition 2.1 (*Asymptotic Validity of Prediction Intervals*)**.** Let L_n, U_n be functions of the data $X_1, \ldots, X_n$. The interval $[L_n, U_n]$ will be called a $(1 - \alpha)100\%$ *asymptotically valid* prediction interval for X_{n+1} given $X_1, \ldots, X_n$ if

$$P(L_n \leq X_{n+1} \leq U_n) \to 1 - \alpha \quad \text{as } n \to \infty \tag{2.2}$$

for all $(X_1, \ldots, X_n)$ in a set that has (unconditional) probability equal to one.

Please cite this article in press as: Pan, L., Politis, D.N., Bootstrap prediction intervals for Markov processes. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.05.010