Universal change point testing for dependent data

Test d'ipotesi universale per punti di cambio in dati dipendenti

Federica Spoto, Alessia Caponera and Pierpaolo Brutti

Abstract One of the main interests in time series analysis is the detection of the so called change-points, defined as timestamps where the model parameters experience a substantial shift in value. Once a candidate change-point is identified, we may want to test whether there is a significant difference in distribution before and after the structural break. In this work we approach the problem from a split-sample perspective and we implement and test on both simulated and real data a two-sample test for time dependent streams that we call *universal change-point testing.*
Abstract *Uno dei problemi principali nell'analisi di serie storiche è l'individuazione dei cosiddetti punti di cambio (change-points), definiti come quegli istanti temporali in cui i parametri del modello vanno incontro ad una sostanziale variazione. Più specificatamente, una volta individuato un valore candidato per il punto di cambio, ci possiamo chiedere se c'è sufficiente evidenza sperimentale a favore dell'esistenza di una differenza significativa nella distribuzione dei dati prima e dopo tale valore. In questo lavoro affrontiamo il problema con delle opportune tecniche di suddivisione del campione, implementando e testando su dati, sia simulati che reali, un test a due campioni per serie dipendenti che chiameremo "test universale per il punto di cambio".*

Key words: Universal Inference, Change-point detection, Likelihood ratio test, Time series model

Federica Spoto
Sapienza University of Rome, e-mail: `federica.spoto@uniroma1.it`

Alessia Caponera
École Polytechnique Fédérale de Lausanne, e-mail: `alessia.caponera@epfl.ch`

Pierpaolo Brutti
Sapienza University of Rome, e-mail: `pierpaolo.brutti@uniroma1.it`

1 Introduction

In temporal processes analysis, one of the main interests is the detection and the analysis of the so called change-points, defined as timestamps where the model parameters have a change in value.

Once a change-point is given as an endogenous information, one may want to test whether there is a significant difference in distribution before and after the structural break. To do so, in this work we approach the problem from the split-sample perspective discussed in [5] (mostly for i.i.d samples) and we implement a two-sample test for time dependent streams of data. We refer to such procedure as *universal change-point testing* and the key steps are formally described in Section 2 for the case of an autoregressive change-point model.

This basic procedure can also be seen as the stepping stone for building a more structured *sequential change-point detection*, which finds important applications for instance in industrial quality control, environment surveillance, computer network security (see among others [4]).

2 Materials and Methods

In this section, we introduce the autoregressive change-point model and the main steps to build the universal change-point testing procedure in this context. For the sake of simplicity, our arguments are presented for a AR(1) model, allowing a single change-point; however, the analysis can be generalized to higher autoregressive orders and multiple change-points. In this setting, the model is written as the composition of two stationary AR segments and takes the form

$$X_t = \begin{cases} \phi_1 X_{t-1} + \varepsilon_{1t} & t \leq \tau \\ \phi_2 X_{t-1} + \varepsilon_{2t} & t > \tau \end{cases}, \tag{1}$$

that, given τ, are assumed to be independent. We consider centered X_t's, but clearly one can allows for changes also in the mean, thus including two different intercepts in the model. We also assume that the errors ε_{jt} are Gaussian with mean zero variance σ_j^2, $j = 1, 2$.

We assume to be able to observe a finite stretch $\{X_1, \ldots, X_N\}$ and to know that a change-point occurred at $1 < \tau < N$. Our goal is to test whether there is a significative difference among the set of parameters before and after the structural break. Hence, formally, define $\theta_1 = (\phi_1, \sigma_1^2), \theta_2 = (\phi_2, \sigma_2^2) \in (-1, 1) \times (0, \infty)$ and consider the following set of hypotheses:

$$\begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 \neq \theta_2 \end{cases}.$$

Then, split $X_1,\dots,X_\tau$ into two segments A_0, A_1, and $X_{\tau+1},\dots,X_N$ into other two segments B_0, B_1, that is,

$$A_1 = (X_1,\dots,X_\gamma), \qquad A_0 = (X_{\gamma+1},\dots,X_\tau),$$
$$B_1 = (X_{\tau+1},\dots,X_{\tau+\gamma}), \qquad B_0 = (X_{\tau+\gamma+1},\dots,X_N),$$

for some integer $1 < \gamma < \tau$. We can hence define $D_0(\tau) = A_0 \cup B_0$, $D_1(\tau) = A_1 \cup B_1$ and the *conditional likelihood*

$$L_{0|1}(\theta_1,\theta_2) = p_{\theta_1,\theta_2}(D_0|D1) = p_{\theta_1,\theta_2}(B_0|B_1)p_{\theta_1,\theta_2}(A_0|A_1).$$

In particular for the autoregressive model (1), we have

$$\begin{aligned} p_{\theta_1}(A_0|A_1) = p_{\theta_1}(X_{\gamma+1},\dots,X_\tau|X_1,\dots,X_\gamma) &= \prod_{t=\gamma+1}^{\tau} p_{\theta_1}(X_t|X_{t-1}) \\ &= \prod_{t=\gamma+1}^{\tau} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2}\frac{(X_t-\phi_1 X_{t-1})^2}{\sigma_1^2}\right\}, \end{aligned}$$

and similarly,

$$\begin{aligned} p_{\theta_2}(B_0|B_1) = p_{\theta_2}(X_{\tau+\gamma+1},\dots,X_N|X_{\tau+1},\dots,X_{\tau+\gamma}) &= \prod_{t=\tau+\gamma+1}^{N} p_{\theta_2}(X_t|X_{t-1}), \\ &= \prod_{t=\tau+\gamma+1}^{N} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2}\frac{(X_t-\phi_2 X_{t-1})^2}{\sigma_2^2}\right\}. \end{aligned}$$

Under H_0, $\theta_1 = \theta_2 = \theta$ and we can compute the MLE of θ based on D_0 as

$$\widehat{\theta} = \arg\max_\theta L_{0|1}(\theta,\theta).$$

Let then $\widehat{\theta}_1, \widehat{\theta}_2$ be any estimators based on D_1 (under the alternative hypothesis).

$$\text{Reject } H_0 \text{ if:} \quad \frac{L_{0|1}(\widehat{\theta}_1,\widehat{\theta}_2)}{L_{0|1}(\widehat{\theta},\widehat{\theta})} > \frac{1}{\alpha}.$$

This is essentially a modified version of the usual likelihood ratio statistic with an *out-of-sample* estimator in the numerator and a “universal” threshold, in the sense that it does not rely on approximations based on limiting distributions, but ensures finite sample guarantees (without additional regularity conditions). Indeed, it is possible to show that the Type-I error can be (conservatively) controlled at level α.

We stress that this method is general and can be applied to different contexts, as long as the conditional likelihood can be computed. The idea is then to extend the result to more general dependency structures, and also to implement a multiple testing procedure that jointly test different values of τ, to then obtain a complete detection procedure.

The present framework could be also generalized to the so-called *spherical functional autoregressive model*, defined for collections of time dependent random fields on a spherical domain (see for instance [1]). In this setting, an alternative change-point analysis is discussed in [3]. All these topics are the object of current ongoing research.

3 Results

We tested the performance of our proposal on synthetic time series and then on a real data-set collecting the *wave heights* in the East Scotian Slope, Canada.

The simulation study has been carried out on two different scenarios to explore *size* and *power* of the universal LRT. In the first scenario we work under the null with $\theta_1 = \theta_2 = (\phi = 0.4, \sigma^2 = 0.3)$, and the observations are generated from the following AR(1) model:

$$X_t = 0.4 \cdot X_{t-1} + \varepsilon_t \qquad 1 \leq t \leq N,$$

where ε_t are Gaussian with zero mean and variance $\sigma^2 = 0.3$. We evaluated the Type I error probability based on $M = 10{,}000$ runs with increasing sample size N and $\alpha = 5\%$. In this scenario $\tau = \frac{N}{2}$ and $\gamma = \frac{\tau}{2}$. The results are shown in Table 1.

N	Test Size
100	0.0009
500	0.0004
1000	0.0001

Table 1: Type I error probability at different sample sizes.

The second scenario follows the alternative hypothesis $\theta_1 \neq \theta_2$. We considered three different settings, increasing the distance between the parameters vectors $\theta_1 = (\phi_1, \sigma_1^2)$ and $\theta_2 = (\phi_2, \sigma_2^2)$. More specifically, we kept the model parameters of the first segment fixed at $\theta_1 = (\phi_1 = 0.4, \sigma_2^2 = 0.2)$ while varying the parameters of the second segment from $\theta_2 = (\phi_2 = 0.6, \sigma_2^2 = 0.3)$ in *Setting 1*, to $\theta_2 = (\phi_2 = 0.8, \sigma_2^2 = 0.4)$ in *Setting 2*, and finally $\theta_2 = (\phi_2 = -0.8, \sigma_1^2 = 0.5)$ in *Setting 3*. For each setting we generated $N \in \{100, 500, 1000\}$ observations with $\tau = \frac{N}{2}$, and $\gamma = \frac{\tau}{2}$. Each setting has been simulated $M = 10{,}000$ times and we evaluated the power of the test with $\alpha = 5\%$. Table 2 contains the results.

As anticipated, we also applied our universal change-point test on the publicly available time series of wave heights collected by Fisheries and Oceans

	N	Power
Setting 1	100	0.010
	500	0.297
	1000	0.696
Setting 2	100	0.117
	500	0.907
	1000	0.996
Setting 3	100	0.781
	500	0.999
	1000	1.000

Table 2: Power of the test for different settings.

Canada (more specifically by the East Scotian Slope buoy), and contained in the `changepoint R` package (see the Github repository). The observations were taken at hourly intervals from January 2005 until September 2012. Here we focus only on the period January - September 2005. The detection of change-points in this time series is helpful to get a better understanding of the variability of the ocean in a certain period of the year, a crucial information for planning operations on offshore infrastructures whose risk of failure strongly increase in the presence of larger wave heights. Quite understandingly, we expect a transition point when moving fro winter to summer, but its exact timing is unknown.

This dataset has been previously used by [2] to test the performance of their algorithm. The analysis there focused on the first order difference of the original data and, between January and September 2005, they detected a change-point at the beginning of April 2005 (dashed blue line in Fig. 1), a result that is consistent with the seasonal behaviour of wave heights. As a confirmatory step, we applied our technique to test the presence of a significant change in the data by setting $\tau = \{\text{April } 1^{\mathtt{st}}\}$. On the log-scale, the rejection rule is then:

$$\text{Reject } H_0 \text{ if:} \quad \log\left(L_{0|1}(\widehat{\theta}_1, \widehat{\theta}_2)\right) - \log\left(L_{0|1}(\widehat{\theta}, \widehat{\theta})\right) > -\log(\alpha).$$

The test confirmed the detection in early April, rejecting the null hypothesis and estimating the model parameters as $\widehat{\theta}_1 = (\widehat{\phi}_1 = 0.012, \widehat{\sigma}_1^2 = 0.048)$, and $\widehat{\theta}_2 = (\widehat{\phi}_2 = 0.029, \widehat{\sigma}_2^2 = 0.019)$

References

1. Caponera, A., Marinucci, D.: Asymptotics for spherical functional autoregressions. Annals of Statistics **49**(1), 346–369 (2021)
2. Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association **107**(500), 1590–1598 (2012). DOI 10.1080/01621459.2012.737745. URL `http://dx.doi.org/10.1080/01621459.2012.737745`

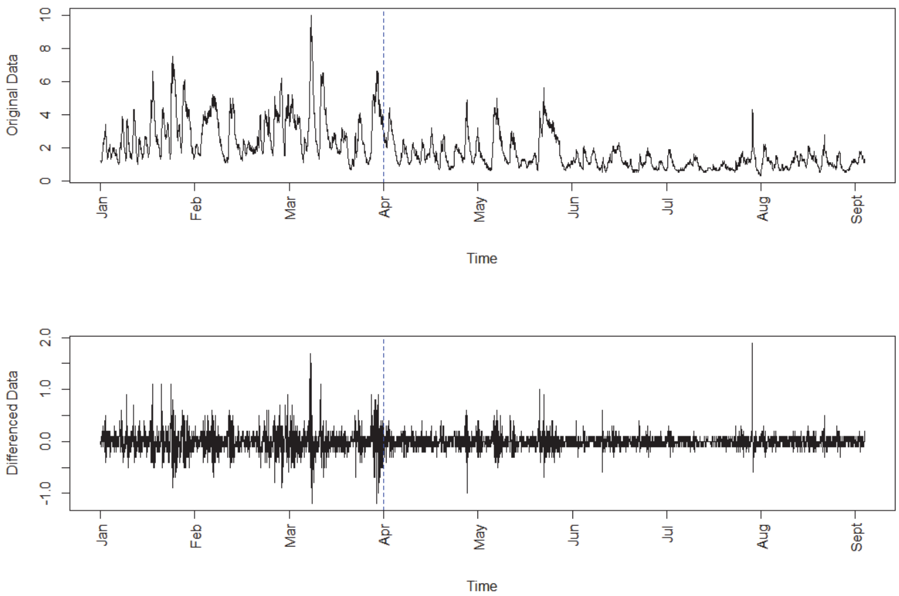

Fig. 1: Top: Original North Atlantic Wave Heights. Bottom: Differenced North Atlantic Wave Heights.

3. Spoto, F., Caponera, A., Brutti, P.: Spherical autoregressive change-point detection with applications. In: Book of Short Papers SIS 2021 (2021). ISBN: 9788891927361
4. Tartakovsky, A.: Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules (1st ed.). Chapman and Hall/CRC (2019)
5. Wasserman, L., Ramdas, A., Balakrishnan, S.: Universal inference. Proceedings of the National Academy of Sciences **117**(29), 16,880–16,890 (2020). DOI 10.1073/pnas.1922664117. URL `https://www.pnas.org/content/117/29/16880`