Available online at www.sciencedirect.com
SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 341 (2005) 627–630

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistics

Nonparametric change-point estimation for dependent sequences

Samir Ben Hariz [a], Jonathan J. Wylie [b], Qiang Zhang [b,1]

[a] *Laboratoire de statistique et processus, département de mathématiques, université du Maine, avenue Olivier-Messiaen, 72085 Le Mans cedex 9, France*

[b] *Department of Mathematics, City University of Hong Kong, 83, Tat Chee Avenue, Kowloon, Hong Kong*

Received 3 June 2005; accepted after revision 27 September 2005

Presented by Philippe G. Ciarlet

Abstract

We present a class of nonparametric change-point estimators for a possibly nonstationary sequence. The estimators are defined using the empirical measures and a semi-norm on the space of measures defined via a family of functions. Using a general setting we prove the rate of $1/n$ convergence in probability. Surprisingly, this optimal rate holds for independent, short-range dependent and long-range dependent sequences. ***To cite this article: S. Ben Hariz et al., C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

Résumé

Estimation non-paramétrique de rupture pour des suites dépendantes. On présente une famille d'estimateurs du temps de rupture dans une suite d'observations non nécessairement stationnaire. Les estimateurs sont définis à partir des mesures empiriques et d'une semi-norme sur l'espace des mesures, définie à l'aide d'une famille de fonctions. Nous montrons alors dans une approche unifiée, que les estimateurs convergent en probabilité avec la vitesse optimale de $1/n$, et ceci aussi bien pour des suites faiblement dépendantes que pour des suites fortement dépendantes. ***Pour citer cet article : S. Ben Hariz et al., C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

1. Introduction

The change-point problem, in which one must detect a change in the marginal distribution of a random sequence, is important in a wide range of applications and has therefore become a classical problem in statistics (Csörgő and Horváth [3]). In this Note we consider the general case of nonparametric estimation that must be used when no a priori information regarding the marginal distributions before and after the change-point is known. We consider this challenging problem and develop a unified framework in which we can deal with sequences with quite general dependence structures. For dependent data one typically expects that the rate of convergence of estimators will become worse as the dependence becomes stronger. However, we prove that the rate of convergence of a broad family of

E-mail addresses: samir.ben_hariz@univ-lemans.fr (S. Ben Hariz), mawylie@cityu.edu.hk (J.J. Wylie), mazq@cityu.edu.hk (Q. Zhang).
[1] Qiang Zhang was supported by City University of Hong Kong grant 7001158.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.09.047

nonparametric estimators is $O_p(n^{-1})$. This is a particularly surprising result because the dependence structure of the sequence plays absolutely no role in determining the rate of convergence.

For independent sequences both parametric and nonparametric methods have been widely studied (Carlstein [2], Dumbgen [4]). However, in recent years the importance of long-range dependent (LRD) processes has been realized in a wide range of applications. We define sequences $(X_i)_{i=1,\ldots,n}$ to be short-range dependent (SRD) if $n^{-1}\mathbb{E}[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])]^2 < \infty$ as $n \to \infty$ and LRD otherwise.

Parametric change-point estimation for LRD sequences, in which one typically has a priori knowledge about the marginal distributions, has been considered by a number of authors (Kokoszka and Leipus [6], Horvath and Kokoszka [5], Ben Hariz and Wylie [1]). Estimating change-points for LRD sequences poses a number of significant challenges and there are much fewer known results in this case.

We adopt a very general framework that allows us to consider a general class of dependence structures. We make no assumption about stationarity in the dependence structure. This is especially important in practice because one can confidently make use of the proposed estimators on a sequence without checking for such stationarity. Our framework represents a unified setting in which independent, SRD and LRD sequences can be treated. We prove the consistency of a Dumbgen-type estimator and show that the $O_p(n^{-1})$ rate of convergence for independent sequences is also achieved for both SRD and LRD sequences.

2. Outline of the results

Let $(X_i)_{i=1,\ldots,n}$ be a real sequence that may be either independent, SRD or LRD. The marginal distribution (which may depend on the sequence length n) is given by

$$\mathcal{L}(X_i) = \begin{cases} P_n & \text{if } i \leqslant n\theta, \\ Q_n & \text{if } i > n\theta, \end{cases}$$

where $0 < \theta < 1$ is the location of the change-point. This means that we assume first-order stationarity (i.e. the marginal distribution is time independent) on either side of the change-point, but make no assumption about stationarity in the dependence structure of the sequence.

Given the sequence, $(X_i)_{i=1,\ldots,n}$, we aim to estimate the location of the change-point θ using an estimator of the following general type:

$$\hat{\theta}_n = \frac{1}{n}\Bigl(\operatorname*{argmax}_{1\leqslant k<n}\bigl\{N(D_k)\bigr\}\Bigr), \tag{1}$$

where N is a (possibly random) semi-norm on the space $\mathcal{M}$ of signed finite measures,

$$D_k = \left[\frac{k}{n}\left(1 - \frac{k}{n}\right)\right]^{1-\gamma}\left(\frac{1}{k}\sum_{i=1}^{k}\delta_{X_i} - \frac{1}{n-k}\sum_{i=k+1}^{n}\delta_{X_i}\right), \tag{2}$$

where δ_{X_i} is the delta measure and γ is a parameter satisfying $0 \leqslant \gamma < 1$. This estimator considers each point as a candidate for the change-point, evaluates the semi-norm and chooses the point with the largest semi-norm as the change-point. In the unlikely event that the maximum value of $N(D_k)$ in (1) occurs at multiple points one can choose any of these points, but for definiteness we take the minimum.

We will develop a framework that can deal with a very general class of estimators. Different semi-norms represent using different measures for the difference between the distributions before and after the change-point. We will show that a very wide class of estimators are appropriate for estimating change-points in dependent data. The semi-norm of a measure ν is defined via $\{\nu(f),\ f \in \mathcal{F}\}$ where $\nu(f) \equiv \int f(x)\nu(\mathrm{d}x)$, and $\mathcal{F}$ is a family of functions.

2.1. Examples of semi-norms

– For $\mathcal{F} = \{\mathbb{1}_{\cdot<X_i},\ i = 1,\ldots,n\}$, we define semi-norms via the quantities $d_i = \nu(\mathbb{1}_{\cdot<X_i})$. This corresponds to the setting of Carlstein [2]. For example, $N(\nu) = \sup_{1\leqslant i\leqslant n}|d_i|$ or $N_p(\nu) = (\frac{1}{n}\sum_{i=1}^n |d_i|^p)^{1/p}$. Observe that in this example the family is random and therefore the semi-norm is also random.

– For $\mathcal{F}=\{f^p\colon x\to x^p,\ p=1,\dots,+\infty\}$ and a measure ν, we define the semi-norm by

$$N(\nu)=\sum_{f\in\mathcal{F}} d(f)\bigl|\nu(f)\bigr|,$$

where $d(f)$ is a sequence of positive weights. This includes the parametric estimators in which we estimate a change in certain moments.
– For $\mathcal{F}=\{\mathbb{1}_{\cdot<x},\ x\in\mathbb{R}\}$ we define N by $N(\nu)=\sup_{f\in\mathcal{F}}|\nu(f)|$. This corresponds to the Kolmogorov–Smirnov norm. Dumbgen [4], considered the family $\mathcal{F}=\{\mathbb{1}_D,\ D\in\mathcal{D}\}$ where $\mathcal{D}$ is a VC subclass, and used semi-norms such that $N(\nu)\leqslant\sup_{f\in\mathcal{F}}|\nu(f)|$.

We now turn our attention to the dependence structure of the sequence. For a given sequence we will allow the estimator to use families of functions that satisfy the following condition.

Assumption 1. There exists constants $C>0$ and $\rho>0$ independent of the sequence length such that

$$\sup_{f\in\mathcal{F}}\sup_{1\leqslant i\leqslant n-m}\bigl|\mathrm{corr}(f(X_i),f(X_{i+m})\bigr|\leqslant Cm^{-\rho}. \tag{3}$$

This assumption simply states that for each of the functions $f\in\mathcal{F}$, the correlation between $f(X_i)$ and $f(X_{i+m})$ must decay algebraically or faster as $m\to\infty$. It is satisfied for a very general class of data. It only excludes the rather artificial case in which the correlation of $f(X_i)$ and $f(X_{i+m})$ decays slower than algebraically.

In Theorem 2.1 we consider semi-norms that are bounded by weighted moments of a countable family of functions and derive conditions under which the optimal convergence rate is achieved.

Theorem 2.1. *Assume that the semi-norm N satisfies $N(\nu)\leqslant\sum_{f\in\mathcal{F}}d(f)|\nu(f)|$, where $\mathcal{F}$ is a countable family of functions satisfying* (3) *and $d(f)$ are positive constants such that $\sum_{f\in\mathcal{F}}d(f)\|f\|<\infty$. If there exists a real number $b>0$, such that*

$$\mathbb{P}\bigl[N(P_n-Q_n)>b\bigr]\to 1\quad \text{as } n\to\infty, \tag{4}$$

then we have $\hat\theta_n-\theta=\mathrm{O}_p(n^{-1})$, where $\hat\theta_n$ is defined in (1) *and $\|f\|=\sup_{n\in\mathbb{N}}(P_n(f^2)+Q_n(f^2))^{1/2}$.*

We now turn our attention to the case in which the family $\mathcal{F}$ contains an uncountable infinity of functions. In this case we need to control the size of the family by using the covering number defined below.

Definition 2.2. Given two functions l and u, the bracket $[l,u]$ is the set of all functions f with $l\leqslant f\leqslant u$. Given a norm $\|\cdot\|$ on a space containing $\mathcal{F}$, an $\varepsilon-$bracket for $\|\cdot\|$ is a bracket $[l,u]$ with $\|l-u\|<\varepsilon$. The bracketing number $N_{[\,]}(\varepsilon,\|\cdot\|,\mathcal{F})$ is the minimal number of ε-brackets needed to cover $\mathcal{F}$.

The following theorem deals with an extremely general set of semi-norms including all of those considered by Carlstein. The theorem essentially states that if the family has a finite bracketing number, we also obtain the $1/n$ rate of convergence.

Theorem 2.3. *Assume that the semi-norm satisfies $N(\nu)\leqslant\sup\{|\nu(f)|,\ f\in\mathcal{F}\}$, where $\mathcal{F}$ is a class of functions satisfying* (3) *and* $\sup\{\|f\|,\ f\in\mathcal{F}\}<\infty$. *We also assume that $\forall\varepsilon>0,\ N_{[\,]}(\varepsilon,\|\cdot\|_X,\mathcal{F})<\infty$, where $\|\cdot\|_X$ is a semi-norm satisfying* $\sup_{n\in\mathbb{N}}|P_n(|f|)|+|Q_n(|f|)|\leqslant\|f\|_X$. *If* (4) *is fulfilled, then we have $\hat\theta_n-\theta=\mathrm{O}_p(n^{-1})$.*

3. Sketch of the proofs

We provide a sketch for the proof of Theorem 2.1. To prove the consistency of the estimator, we define $D_n(t)\equiv D_{[nt]}$ and decompose $D_n(t)$ into a mean part and a centered random part $D_n(t)=(P_n-Q_n)h(t)+B_n^w(t)$, where $h(t)=t^{-\gamma}(1-t)^{-\gamma}(t(1-\theta)\mathbb{1}_{t\leqslant\theta}+\theta(1-t)\mathbb{1}_{t>\theta})$. Then, for any $\eta>0$, we show

$$\mathbb{P}\bigl[|\hat\theta_n-\theta|>\eta\bigr]\leqslant\mathbb{P}\left[N\bigl(B_n^w(\hat\theta_n)\bigr)\geqslant\frac{ab}{2},\ |\hat\theta_n-\theta|>\eta\right]+\mathbb{P}\left[N\bigl(B_n^w(\theta)\bigr)\geqslant\frac{ab}{2}\right]+\mathbb{P}\bigl[N(P_n-Q_n)>b\bigr]. \tag{5}$$

We control the first term by decomposing the set $\{t\colon |t-\theta|>\eta\}$ into shells

$$S_j=\bigl\{t\colon 2^{-j}<t(\theta-\eta)^{-1}\leqslant 2^{-j+1}\bigr\}\cup\bigl\{2^{-j}<(1-t)(1-\theta-\eta)^{-1}\leqslant 2^{-j+1}\bigr\}.$$

We then use the following maximal inequality from Moricz [7]: there exists a constant $D(\rho)$ such that

$$\mathbb{E}\Biggl(\max_{1\leqslant k\leqslant n}\Biggl|\sum_{i=1}^{k}f(X_i)-\mathbb{E}\bigl(f(X_i)\bigr)\Biggr|\Biggr)^2\leqslant D^2(\rho)\|f\|^2n^{2-\rho}. \tag{6}$$

We control the second term using (6) and the third term tends to zero by assumption.

For the rate proof we first show that, for any positive integer M, positive real numbers b and c and positive sequence r_n, we have $\mathbb{P}(r_n|\hat\theta_n-\theta|>2^M)\leqslant E_1+E_2+E_3$, where $E_1\equiv\mathbb{P}(N(B_n^w(\theta))>c)$, $E_2\equiv\mathbb{P}[r_n|\hat\theta_n-\theta|>2^M,\ N(B_n^w(\hat\theta_n)-B_n^w(\theta))\geqslant C_h(bh(\theta)-2c)|\hat\theta_n-\theta|]$, $E_3\equiv\mathbb{P}(N(\delta_n)\leqslant b)$ and C_h is a constant that only depends on θ and γ. We then bound E_2 using a decomposition of the set $\{t\colon r_n|t-\theta|>2^M\}$ into shells, $S_{n,j}=\{t\colon 2^j<r_n|t-\theta|\leqslant 2^{j+1}\}$. The consistency result (5) and the following lemma which controls the size of oscillations of $B_n^w(t)$ complete the proof.

Lemma 3.1. *Assume* (3) *with* $\rho<1$, *then there exist constants* $C(\theta,\eta)$ *and* $D(\rho)$ *such that for* $\kappa<\eta$,

$$\mathbb{E}\Bigl(\sup_{|t-\theta|\leqslant\kappa}\bigl|\bigl(B_n^w(t)-B_n^w(\theta)\bigr)(f)\bigr|\Bigr)\leqslant C(\theta,\eta)D(\rho)\|f\|n^{-\rho/2}\kappa^{1-\rho/2}.$$

To prove Theorem 2.3, we need a projection argument to deal with the uncountable family of functions. Although the proof is more tedious, the procedure is similar to that for Theorem 2.1. We end this outline by noting that the proof can easily be adapted to handle vector-valued sequences.

References

[1] S. Ben Hariz, J.J. Wylie, Rates of convergence for the change-point estimator for long-range dependent sequences, Probab. Statist. Lett. (2005), in press.
[2] E. Carlstein, Nonparametric change-point estimation, Ann. Statist. 16 (1988) 188–197.
[3] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, Wiley, Chichester, 1997.
[4] L. Dumbgen, The asymptotic behavior of some nonparametric change-point estimators, Ann. Statist. 19 (1991) 1471–1495.
[5] L. Horváth, P. Kokoszka, The effect of long-range dependence on change-point estimators, J. Statist. Plann. Inference 64 (1997) 57–81.
[6] P. Kokoszka, R. Leipus, Change-point in the mean of dependent observations, Statist. Probab. Lett. 40 (1998) 385–393.
[7] F. Moricz, Moment inequalities and the strong laws of large numbers, Z. Wahr. Verw. Geb. 35 (4) (1976) 299–314.