Journal of Statistical Planning and Inference 193 (2018) 151–178

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Optimal change point detection in Gaussian processes[☆]

Hossein Keshavarz [a,*], Clayton Scott [b,a], XuanLong Nguyen [a,b]

[a] *Department of Statistics, University of Michigan, United States*
[b] *Department of Electrical Engineering and Computer Science, University of Michigan, United States*

ARTICLE INFO

Article history:
Received 10 May 2017
Received in revised form 7 September 2017
Accepted 9 September 2017
Available online 22 September 2017

Keywords:
Change-point detection
Gaussian processes
Fixed domain asymptotic analysis
Minimax optimality

ABSTRACT

We study the problem of detecting a change in the mean of one-dimensional Gaussian process data in the fixed domain regime. We propose a detection procedure based on the generalized likelihood ratio test (GLRT), and show that our method achieves asymptotically near-optimal rate in a minimax sense. The notable feature of the proposed method is that it exploits in an efficient way the data dependence captured by the Gaussian process covariance structure. When the covariance is not known, we propose the plug-in GLRT method and derive conditions under which the method remains asymptotically near-optimal. By contrast, the standard CUSUM method, which does not account for the covariance structure, is shown to be suboptimal. Our algorithms and asymptotic analysis are applicable to a number of covariance structures, including the Matern class, the powered exponential class, and others. The plug-in GLRT method is shown to perform well for maximum likelihood estimators with a dense covariance matrix.

1. Introduction

Change point detection is the problem of detecting an abrupt change or changes arising in a sequence of observed samples. A common problem of this type involves detecting shifts in the mean of a temporal or spatial process. This problem has found a variety of applications in many fields, including audio analysis (Gillet et al., 2007), EEG segmentation (Lavielle, 2005), structural health monitoring (Noh et al., 2012; Hu et al., 2007) and environment sciences (Last and Shumway, 2008; Verbesselt et al., 2010). Despite advances in the development of algorithms (Kawahara and Sugiyama, 2009; Lavielle, 2005; Liu et al., 2010; Rigaill, 2010) and asymptotic theory (Bertrand et al., 2011; Tartakovsky et al., 2006; Shao and Zhang, 2010; Levy-leduc, 2007) for a number of contexts, such studies are mainly confined to the setting of (conditionally) independently distributed data. Existing works on optimal detection of shifts in the mean in temporal data with statistically dependent observations are far less common.

Incorporating dependence structures into the modeling of random processes is a natural approach. In fact, this has been considered in detecting changes of remotely collected data (Chandola and Vatsavai, 2011; Gabriel et al., 2011). For instance, Chandola and Vatsavai (2011) proposed a Gaussian process based algorithm to identify changes in Normalized Difference Vegetation Index (NDVI) time series for a particular location in California. It is therefore of interest to study how the dependence structures of the underlying process can be accounted for, e.g., its covariance function and spectral density, in designing statistically efficient detection procedures. In this paper, we shall focus on the detection of a single change in the mean of a Gaussian process data sequence.

[☆] This research is partially supported by National Science Foundation (NSF) grant ACI-1047871. Additionally, CS is partially supported by NSF grants CCF-1422157, CCF-1217880, and CCF-0953135, and LN by NSF CAREER award DMS-1351362, NSF CNS-1409303, and NSF CCF-1115769.

* Corresponding author.
E-mail addresses: hksh@umich.edu (H. Keshavarz), clayscot@umich.edu (C. Scott), xuanlong@umich.edu (X. Nguyen).

https://doi.org/10.1016/j.jspi.2017.09.003
0378-3758/© 2017 Elsevier B.V. All rights reserved.

Consider a simplified setting in which we let G be a Gaussian process on a domain $\mathcal{D} \subseteq \mathbb{R}$ and $\mathcal{D}_n := \{t_k\}_{k=1}^n \subset \mathcal{D}$ denote a finite index set of sampling points. Denote the observed samples by $\boldsymbol{X} = \{X_k\}_{k=1}^n$ in which $X_k = G(t_k)$ for $k = 1, \ldots, n$. Moreover, let $t \in \mathcal{C}_{n,\alpha} \subseteq \{1, \ldots, n\}$ (the parameter α is a positive scalar which will be introduced in Section 2.1) and $b > 0$ represents the point of sudden change and the jump/shift value, respectively. Namely, there is $\mu \in \mathbb{R}$ (which will be assumed to be 0 for now) such that

$$\mathbb{E}X_k = \left(\mu - \frac{b}{2}\right) \mathbb{1}(k < t) + \left(\mu + \frac{b}{2}\right) \mathbb{1}(k \geq t), \quad k \in \{1, \ldots, n\}. \tag{1.1}$$

To design a detection procedure and analyze its performance as sample size n grows to infinity, one is confronted with two fundamentally different frameworks, the framework of *increasing domain asymptotics* and that of *fixed domain (infill) asymptotics*, cf., e.g., Rao et al. (2012). The former arises naturally in time series analysis, which is distinguished by the constraint that the distance between consecutive sampling *time* points are bounded away from zero. The simplest instance of the sampling scenario in this regime arises when the diameter of $\mathcal{D}_n$ is of order n and $\min |t_{i+1} - t_i| > \epsilon$ for some strictly positive, fixed scalar ϵ. In our notation the index set for the Gaussian process represents the sampling time points. Typically we set $\mathcal{D} = \mathbb{R}$ and $\bigcup_{n=1}^\infty \mathcal{D}_n = \mathbb{N}$ or $\mathbb{Z}$. There is a large literature on change point detection via the increasing domain asymptotics (Antoch et al., 1997; Horváth, 1997; Horváth and Hušková, 2012; Kokoszka and Leipus, 1998; Rencova, 2009; Yao and Davis, 1986) — which we shall return to in a moment. Fixed domain (or infill) asymptotics, one the other hand, is a more suitable setting when the index set of sampling points $\mathcal{D}$ is bounded, so that the observations get denser in $\mathcal{D}$ as n increases. Particularly for $\mathcal{D} \subset \mathbb{R}$, we have that $\min |t_{i+k} - t_i| = \mathcal{O}(k/n)$ for positive integers i, k with $i, (i+k) \in \{1, \ldots, n\}$, and it can be extended to multidimensional domains in a straightforward way. This is the case for spatially distributed data (Stein, 1999), where the domain of the index set is typically of one, two or three dimensions. This approach is also appropriate in the context of change detection for non-stationary processes (Adak, 1998; Dahlhaus, 1997; Adak, 1998; Last and Shumway, 2008). The development of detection algorithms and theory for fixed domain asymptotics are relatively rare.

To gain a quick intuition on how the different asymptotic settings can affect the detection of a change in the observed sequence $\boldsymbol{X} = \{X_k\}_{k=1}^n$, one can look into the correlation among nearby samples in the sequence. In the increasing domain regime, even for long range dependent processes the correlation among samples X_i and X_j is small when $|j - i|$ is large. By contrast, in the fixed domain regime, regardless of how large the sample size is, if $|j - i|$ is of order n^β for some $\beta \in (0, 1)$, the correlation among X_i and X_j is still close to one. This entails that the effective sample size is much smaller than n. As a consequence, standard techniques that work well in the increasing domain setting do not work as well in the fixed domain setting. In the latter, we shall need more effective techniques to account for the strong dependence in the observed samples.

Previous works. An early attempt to study shift in mean detection was that of Chernoff and Zacks (1964). More general settings of this problem have been studied in subsequent works, e.g., MacNeill (1974), Deshayes and Picard (1985) and Yao and Davis (1986). For instance it is assumed in Yao and Davis (1986) that the sequence of X_k's are independent Gaussian variables. They proposed a detection method based on the generalized likelihood ratio test (GLRT), also known as the *cumulative sum (CUSUM)* test, and given by

$$T_{\text{CUSUM}} = \mathbb{1}\left\{\max_{t \in \mathcal{C}_{n,\alpha}} \left\{\sqrt{\frac{t(n-t)}{n}} \left|\frac{1}{n-t} \sum_{k=t+1}^n X_k - \frac{1}{t} \sum_{k=1}^t X_k\right|\right\} \geq R_n\right\}. \tag{1.2}$$

CUSUM compares the maximum of a test statistic over $C_{n,\alpha}$ with a critical value R_n. Non-asymptotic upper bounds on the error probabilities of this simple test were obtained by the authors under the Gaussian and i.i.d. assumptions. Due to its simplicity, the CUSUM test is very popular, and has been applied to a variety of settings.

For example, subsequent works studied the behavior of the CUSUM test under weaker assumptions in the increasing domain regime (Antoch et al., 1997; Horváth, 1997; Horváth and Hušková, 2012; Rencova, 2009). We wish to mention Rencova ((Rencova, 2009), chapter 4), who studied the same test as Yao and Davis, (1986), but working with the assumption that $\boldsymbol{X}$ is a strong mixing time series. Kokoszka and Leipus (1998) also analyzed the CUSUM test, but working with a different dependent observation model with sub-squared growth of the variance of partial sums, i.e., there is $\delta \in (0, 2)$ such that for any $k < m$, $\text{var}\sum_{j=k}^m X_j \lesssim (m - k + 1)^\delta$. Horváth (1997), Horváth and Hušková (2012) and Antoch et al. (1997) studied the performance of the CUSUM test for the detection of a sudden change in the mean in linear processes, i.e. $X_t = \sum_{j=0}^\infty w_j \epsilon_{t-j}$, in which $\{\epsilon_t\}_{t=-\infty}^\infty$ are i.i.d. and zero mean random variables and the weights $\left\{w_j\right\}_{j=0}^\infty$ satisfy some properties such as absolute or square summability. We also refer the reader to Aue and Horváth (2013) and Horváth and Rice (2014) for a comprehensive review of abrupt change detection in the increasing domain regime.

The CUSUM test may also be applied to one dimensional processes with correlated samples, after a proper standardization. For instance, Horváth and Hušková (2012) used a different normalizing factor for applying CUSUM to one dimensional Gaussian time series with long range dependence. However apart from the standardizing factor, they do not directly incorporate the correlation structures of the data in the formulation of the test statistics. Furthermore, different forms of the CUSUM test were proposed to detect abrupt changes in the sequential detection literature, see e.g., Lai (1998). At first sight, it may seem puzzling how the CUSUM test attains nearly optimal detection performance in the increasing domain even as its test statistic apparently ignore the dependence among data samples (see e.g. Antoch et al. (1997); Horváth (1997); Horváth and Hušková (2012); Rencova (2009)). As noted earlier, the covariance $\text{cov}(X_s, X_t) \to 0$ as $|t - s|$ grows to infinity.