

Kernel Based Approaches for Change-Point  
Detection  
Midway report

Anirudhan J. Rajagopalan, ajr619

April 3, 2016

## Abstract

Electroencephalography (EEG) signals are detected by using electrodes that are applied to the surface of mammals. A sensitive instrument can show continuous fluctuations of the electric potential between the electrodes. These potentials are due to the superposition of the electrical activity of tens of thousands of neuron cells lying on the surface areas of the brain [1]. A typical human EEG signal is produced by 36 electrodes placed on the scalp. This work aims to exploit the spatial relation of the electrodes to identify a Reproducing Kernel Hilbert Space (RKHS) kernel function that will help us find the change points in the EEG signal data.

# 1 Change point detection

## 1.1 Stationary process

A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time. Consequently, parameters that define the process such as mean ( $\mu$ ), standard deviation ( $\sigma$ ) remain constant

## 1.2 Change points

A change point in a time series is the times at which the probability distribution of the time series changes. A change point detection problem involves detecting whether or not a change has occurred, and identifying the times at which such changes occurred.

Change points can be identified when the parameters defining the stationary process, such as mean ( $\mu$ ) and standard deviation ( $\sigma$ ) changes. We discuss ways to identify change points when mean of the process changes.

## 1.3 Offline Change point detection

Offline change point detection or retrospective change point detection methods has the complete data available at the beginning of the program.

### 1.3.1 Univariate Time series

Lets assume a time series of observations  $x_1, x_2, \dots, x_n$  of independent random variables with parameters  $(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2), \dots, (\mu_n, \sigma_n^2)$ . Also lets assume that each of the observation  $x_i$  is normally distributed with mean  $\mu$  and common variance  $\sigma^2 \forall i \in 1, 2, \dots, n$ . When there is no change in mean, the hypothesis of stability (null hypothesis) is defined as

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu \quad (1)$$

Lets suppose that there is a change in the mean in the observations at an unknown point  $K$ . This can be define dy

$$H_1 : \mu_1 = \dots \mu_k \neq \mu_{k+1} \dots = \mu_n \quad (2)$$

In our experiments we are going to assume that we know  $\mu_1, \mu_n$  and  $\sigma$  are known beforehand (Refer 2.1.1 of [2]).

We can find the change poing by using Maximum Likelihood estimates.

Under  $H_0$ ,

$$L_0(\mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2}} \quad (3)$$

$$(4)$$

and the Maximum likelihood estimator is given by

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5)$$

Under  $H_1$ ,

$$L_1(\mu_1, \mu_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\left(\frac{\sum_{i=1}^k (x_i - \mu_1)^2 + \sum_{i=k+1}^n (x_i - \mu_n)^2}{2}\right)} \quad (6)$$

$$(7)$$

and the Maximum likelihood estimator is given by

$$\hat{\mu}_1 = \bar{x}_k = \frac{1}{k} \sum_{i=1}^k x_i \quad (8)$$

$$\hat{\mu}_n = \bar{x}_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^n x_i \quad (9)$$

We can use the Maximum likelihood estimator directly to find the change points in the given data. But calculating the MLE is computationally intractable. An alternate set of equations is given in Chapter 2 of [2]. We use the alternate set of equations to find the change points. The experiments using sample data and its results are shown below.

### 1.3.2 Experiments — Univariate time series

The offline changepoint detection problem, gives a pretty accurate value for changepoint at  $k = 300$ . The different plots are as displayed below.

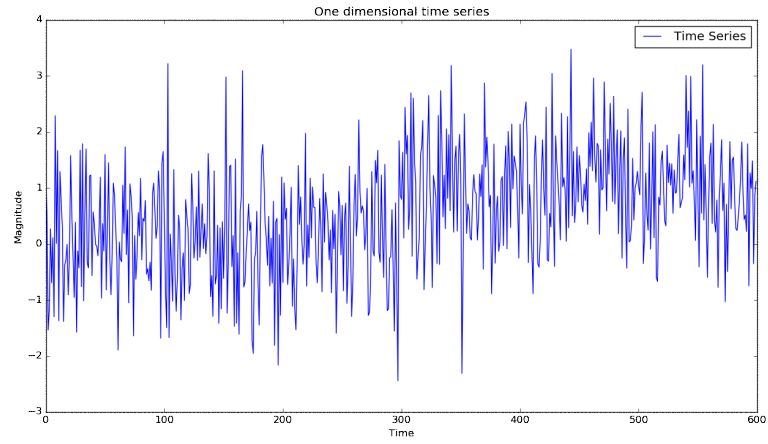


Figure 1: One dimensional time series.

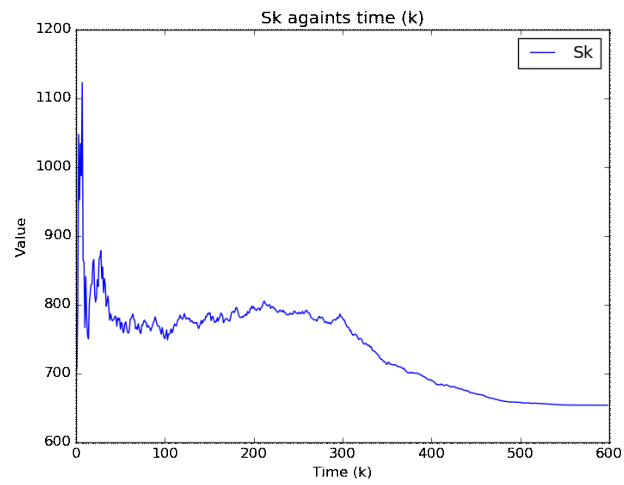


Figure 2: SK values for one dimensional offline detection problem.

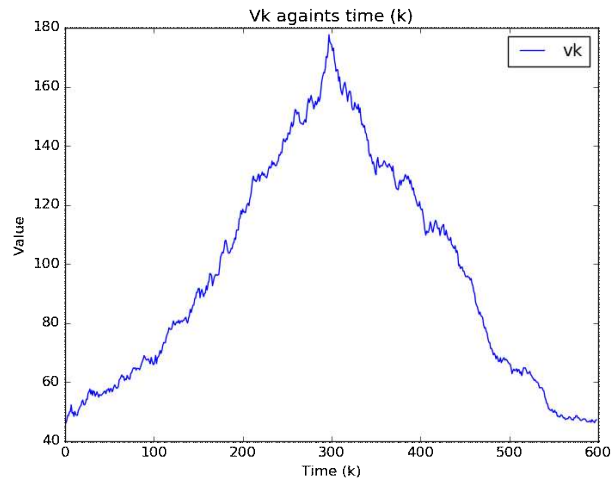


Figure 3: VK values for one dimensional offline detection problem.

### 1.3.3 Multivariate Time series

### 1.4 Online Change point detection

## 2 Spectral Density

### 2.1 Auto CoVariance Function

### 2.2 Auto Corelation Function

### 2.3 Peridogram

### 2.4 Change points using spectral density

## References

- [1] B.E. Brodsky and B.S. Darkhovsky. *Non-Parametric Stastical Diagnosis — Problems and methods*. Springer — Science + Business media, B.V., 2000.
- [2] Jie Chen and Arjun K. Gupta. *Parametric Stastical Change Point Analysis - With applications to Genetics, Medicine and Finance*. Birkhauser, 2012.