Kernel Based Approaches for Change-Point Detection — Report 1

Anirudhan J. Rajagopalan, ajr619

February 18, 2016

1 Offline Detection

1.1 Univariate Detection

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

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \tag{1}$$

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

$$H_1 : \mu_1 = \ldots \mu_k \neq \mu_{k+1} \ldots = \mu_n \tag{2}$$

In our experiments we are going to assume that we know μ_1, μ_n and σ are known beforehand (Refer 2.1.1 of [2]).

1.1.1 Experiments

Finding the likelihood directly using the likelihood function is not practical as it is not computationally tractable for even a small value of n (600 in our case). So we follow the steps given in the reference[2] to find the change point.

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

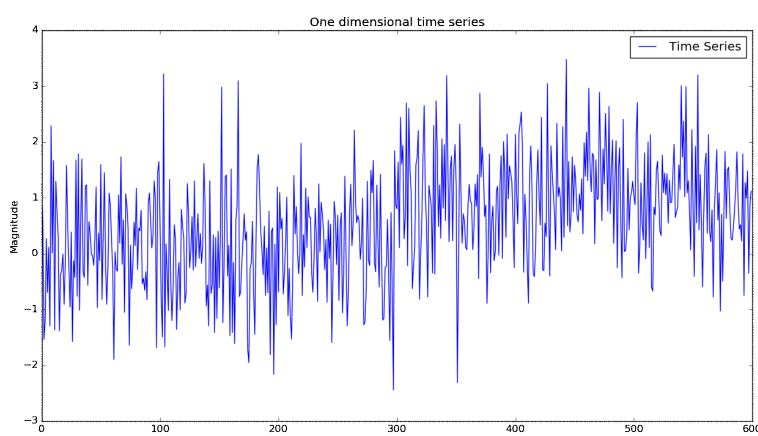

Figure 1: One dimensional time series.

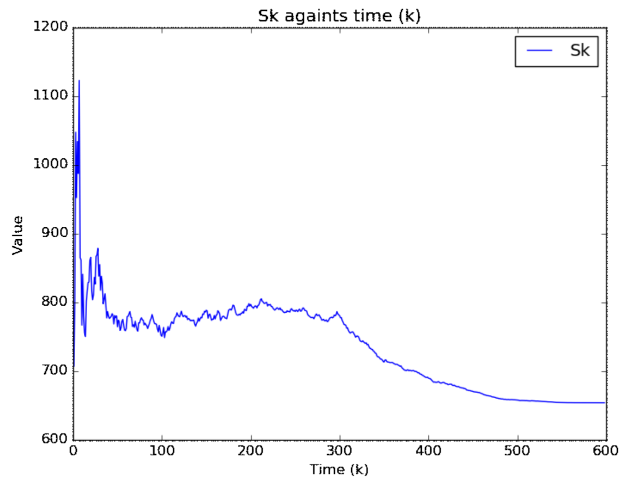

Figure 2: SK values for one dimensional offline detection problem.

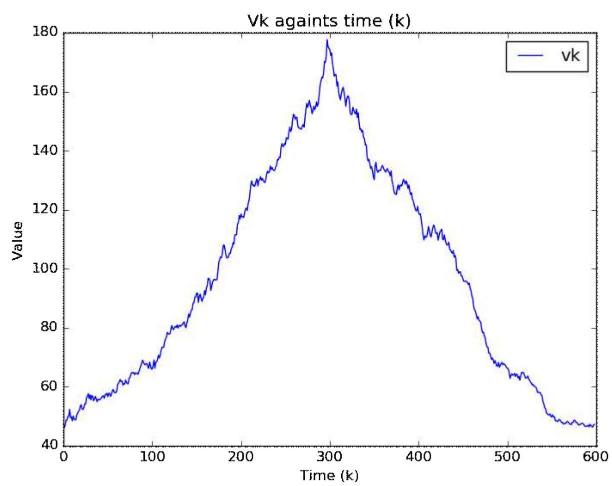

Figure 3: VK values for one dimensional offline detection problem.

1.1.2 Questions & Challenges

1. Finding how to derive sk, vk and T^2 test stastic for our own dataset when mean, and standard deviations are not known.
2. Understanding what sk and vk represent.

1.2 Multi-variate Detection

Multivariate model is similar to the above Univariate model except that each and every observation is m-dimensional. Let $x_1, x_2, \ldots, x_n$ be a sequence of independent m-dimensional normal random vectors with parameters $(\mu_1, \Sigma_1), (\mu_2, \Sigma_2), \ldots, (\mu_n, \Sigma_n)$, respectively. Assume $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_n = \Sigma$. The null hypothesis is given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu(\text{unknown}) \tag{3}$$

If we assume that there is a change at point k in the paramters governing the observation, then the hypothesis (alternate hypothesis) is given by

$$H_1 : \mu_1 = \cdots = \mu_k \neq \mu_{k+1} = \mu_n \tag{4}$$

Where k represents the position of the single change point. (Refer 3.1 of [2]).

This can be solved by following the steps described in section 3.1.1 of [2]. As in the case of univariate model, finding the likelihood directly is computationally intractable as 2π and e has negative powers that can go pretty large and hence the likelihood will always become 1.

1.2.1 Experiments

We did Several experiments by varying the number of dimension and also the total number of samples. So far, in all variations, we are able to identify the change point pretty accurately using the offline detection method described above. The plots below are for dimension = 6 and number of samples = 600 with change point at 300.

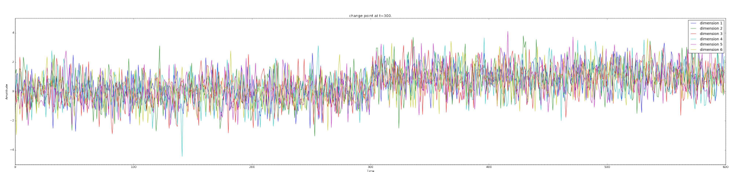

Figure 4: Multi dimensional time series.

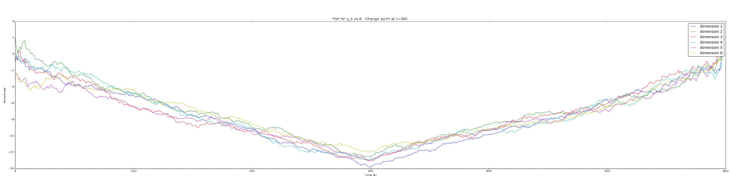

Figure 5: Value of y_k with respect to various k.

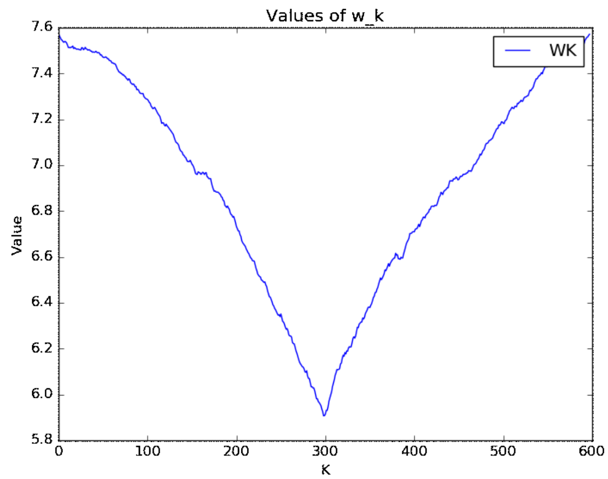

Figure 6: Value of w__k with respect to various k.

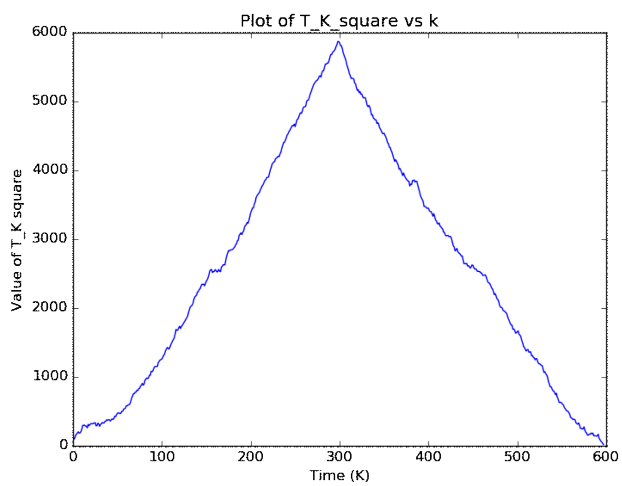

Figure 7: Value of the test stastic T_k^2 with respect to various k.

1.2.2 Questions & Challenges

1. Understanding what the variables yk, wk, and tk__sq represent.
2. Figuring out the reason for the difference between the graphs between one dimensional and multi dimensional change point detection.

2 Online Detection

2.1 Univariate detection

Online detection of change point is slightly difficult than offline detection. We have a series of observation defined by $x_1, x_2, \ldots, x_n$. Each observation denoted by x_i is assumed to be made of N samples.

We start by defining a sufficient stastic as:

$$s_i = \ln \frac{p\theta_1(y_i)}{p\theta_0(y_i)} \tag{5}$$

We also define a decision rule d given by

$$d = \begin{cases} 0 & if S_1^N < h; H_0 \text{ is chosen} \\ 1 & if S_1^N \geq h; H_1 \text{ is chosen} \end{cases} \tag{6}$$

Where h is a threshold chosen by the user. S_1^N is called as the decision function.

We also have a stopping rule which is defined by

$$t_a = N.\min K : d_K = 1 \tag{7}$$

Where d is the decision taken with the aid of the decision function defined above. This method of finding the stopping rule and alarm time is as given in[1]

2.1.1 Experiments

We run experiments using grid search with various values of h and kappa. In this particular experiment, we used values ranging from 1.5 to 5.5 with step size 0.25 for h and 0.5 to 4 with step size of 0.25 for kappa. We then proceed to find all the values of the stopping rule and also the alarm time for h and kappa combination. The time series, the plots for alarm time and stopping rule are plotted below.

2.1.2 Observations

On retrospect, it doesn't make sense to do a grid search over h and kappa values as they dont impact alarm time or the stopping rule jointly. The graphs obtained by plotting h values, kappa values and alarm time/stopping rule reflects this.

2.1.3 Questions & Challenges

1. How to find the correct value for h and kappa given a distribution.

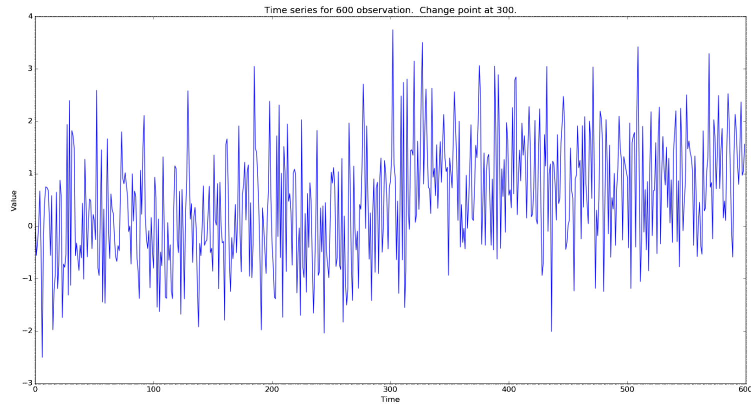

Figure 8: One dimensional time series for online detection.

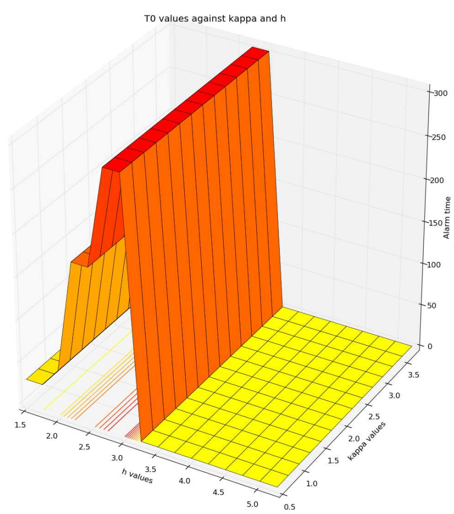

Figure 9: Stopping rule for various values of h and kappa.

References

[1] Michele Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. PTR Prentice-Hall, Inc., 1993.

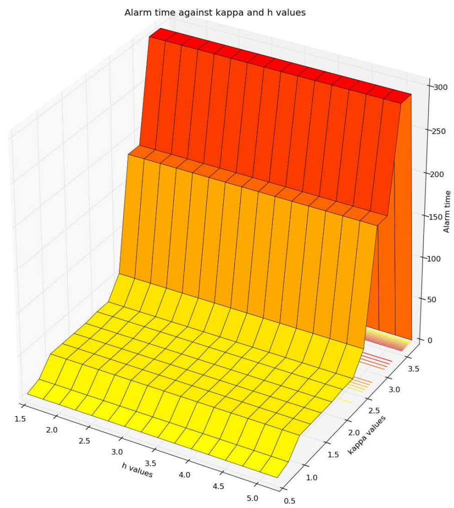

Figure 10: Alarm time for values of h and kappa.

[2] Jie Chen and Arjun K. Gupta. *Parametric Stastical Change Point Analysis - With applications to Genetics, Medicine and Finance*. Birkhauser, 2012.