

THESIS TITLE - CHANGE POINT DETECTION

TYLER MANNING-DAHAN

A THESIS
IN
THE DEPARTMENT
OF
ENGINEERING AND COMPUTER SCIENCE

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF APPLIED SCIENCE
CONCORDIA UNIVERSITY
MONTRÉAL, QUÉBEC, CANADA

MAY 2019

© TYLER MANNING-DAHAN, 2019

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Tyler Manning-Dahan**

Entitled: **Thesis Title - Change Point Detection**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____	Chair
_____	Examiner
_____	Examiner
_____	Examiner
_____	Supervisor

Approved _____
Chair of Department or Graduate Program Director

_____ 20 _____

Abstract

Thesis Title - Change Point Detection

Tyler Manning-Dahan

Text of abstract.

Acknowledgments

I would like to thank my supervisor Dr. Jia Yuan Yu for accepting me into his lab and pushing the limit of my intellectual understanding to places I have never thought possible.

I would also like to thank all my colleagues at DRW who have helped me learn about the finance industry over the past two years. I am especially grateful to Yves, Neil and Laura, who have been very patient with me and have me taught me a lot about understanding financial market micro-structure, writing clean code and building robust statistical models.

Lastly, I would like to thank my fiancée, Tanya, who initially pushed me to go back to school while I was still young and had the opportunity.

Contents

List of Figures	vi
List of Tables	vii
1 Introduction	1
1.1 Motivation	1
1.1.1 Health Care	1
1.1.2 Financial Applications	2
1.2 Characteristics of the change point problem	2
1.3 Problem Formulation	4
1.3.1 Performance Measures	4
1.4 Related Work	5
1.5 Our Contributions	6
2 Chapter 2	7
2.1 Chapter 2 Section	7
2.1.1 Chapter 2 Subsection	7

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Almost 100 years ago, Walter Shewart recognized the need for monitoring manufacturing processes in real-time in a systematic way. The use of statistical control charts allowed for real-time detection of changes in variation that may have indicated a degradation in quality in the production process [15]. Shewart's method was one of the first formal frameworks to solve the problem of detecting changes in a distribution of a sequence of random variables. This problem would come to be known more generally as the *change point detection problem* and would come to apply across various industries. The following are a few motivating examples.

1.1.1 Health Care

Health care is an important area for quickly detecting signal changes in heart rate monitoring [19] [16], epilepsy signal segmentation [11], and multi-modal MRI lesion detection [3] to name a few. Quickly detecting changes to a patient's health is absolutely necessary for any system to be of practical use. However, this quick detection must be balanced with high accuracy as false positives or missed detections could have life-threatening consequences.

1.1.2 Financial Applications

The application of accurate and timely change point detection is also very appealing to the finance sector where shifts in asset prices can happen suddenly. Change point detection is further complicated in financial applications because of the non-stationarities typically observed in asset pricing along with long memory processes. Note, in the financial literature, change-points are also referred to as structural breaks, but for this thesis will use the broader term change-points.

An on-line, quick detection technique is proposed in [14], where a modified Shiryaev - Roberts procedure is used in a case study to detect a change-point on a single stock's daily returns. They compare their non-parametric method with other classic control chart methods using speed of detection and false alarm rate as measures of performance.

Detecting changes in variance is specifically explored in [8]. The authors propose an off-line change-point algorithm that minimizes a global cost function by using an adaptive regularization function. The algorithm is applied to the absolute returns of the FTSE 100 stock index and the US dollar-Japanese Yen foreign intra-day exchange rate to detect changes in asset price volatility. The change-points identified in the FTSE 100 coincided with key market events such as the stock market crash that occurred on October 14th, 1987 and breaking the 5000 price barrier in August 1997.

See section 1.3.6 of [17] for more applications to options markets and arbitrage opportunities.

1.2 Characteristics of the change point problem

A number of surveys of the literature already exist [1] [12], therefore we will not cover all existing methods but rather touch upon several, important factors to consider when tackling the change point detection problem. Across the body of literature, these factors determine what methods are available to practitioners.

The first factor is selecting between *parametric* and *non-parametric* techniques. Deciding between these two broad techniques is dependent on the prior knowledge one wants to encode into the problem. For example, if it is known that data is generated by a distribution from the exponential family of distributions, then we can subset the

problem from the space of all possible distributions to a smaller space of distributions. For example, Shewart control charts and CUSUM change-point techniques are all parametric techniques based on Gaussian-family of distributions [13] [5]. In certain settings, it is not possible to leverage information about the data and non-parametric techniques must be used instead [4].

The second factor is to deciding whether change-points should be detected *offline* or *online*. Some algorithms are off-line—also referred to as batch algorithms or retrospective or a posteriori change-point detection—and they are applied in an ex-post fashion after the dataset has been completely acquired [18]. The aforementioned Shewart control chart and CUSUM algorithm were both designed for data that is streamed in a real-time fashion. In the literature, on-line methods of change point detection are also referred to sequential change point detection [17]. For this thesis, we will use the terms interchangeably.

The third factor is determining if there are multiple change points or to assume there is only a single change point to detect. This is an important factor for off-line change point detection where the decision to detect 1 or more change-points is chosen at the outset. Detecting multiple change-points could also be relevant for the on-line case if a situation arises where the window of time series under consideration may contain more than one change point. However, most on-line change-point methods are designed to detect a single change-point at a time.

Finally, the last factor to address is determining exactly what kinds of statistical changes the algorithm should detect. Many methods focus solely on detecting changes in the mean of a distribution [9]. Some methods are more general and can detect changes in the variance or higher order moments and do not focus on any particular one. Methods like kernel change point detection can typically detect any distributional changes.

This thesis will concern itself with on-line change point detection, where data is received in a streaming nature. We assume no prior distributional characteristics on the data and operate in a completely non-parametric setting.

1.3 Problem Formulation

The basic change-point problem is set up as hypothesis test between two segments of a time series. Since we are concerned with the on-line setting, we will always consider a time series of fixed size, n . Let X_1, \dots, X_n be a series of independent random variables observed. Then, one of the following hypotheses holds:

$$\begin{cases} H_0 : X_1, X_2, \dots, X_n \sim F_0 & \text{(no change-point occurred)} \\ H_1 : X_i \sim F_1, X_j \sim F_2 & \text{(a change-point occurred).} \end{cases} \quad (1)$$

Where $i = 1, 2, \dots, t-1$ and $j = t, \dots, n$ are two distinct segments separated by change-point t that is within the time series window. Furthermore, F_0, F_1, F_2 are cumulative distribution functions (CDFs) with corresponding probability density functions (PDFs), f_0, f_1, f_2 . Because we are operating in a non-parametric setting, the CDFs are assumed to be completely unknown.

Many change-point detection algorithms define a statistic that is computed using each set before and after the possible change-point, t . If the statistic is above a particular threshold then time t is classified as a change-point, $\hat{\tau}$.

In the on-line scenario, the time series under consideration can be thought of as a sliding window with data constantly coming in and out of the window of interest. The size of the window is an important consideration that is typically chosen based on the problem being solved. Too small a window and the sets of data may not yield a statistically significant result. Too large of a window and the problem leans more towards an off-line model, where high volumes must be stored and several change-points may appear in a given window.

1.3.1 Performance Measures

Because of the unsupervised nature of detecting change-points, it is difficult to evaluate the performance of change-point detection models with real world data. Many papers detail asymptotic and non-asymptotic theoretical guarantees of their proposed change-point methods. These theoretical optimalities are typically compared across different change-point methods for benchmarking a new algorithm.

Average run length is often used to compare A and B by Y. It is a function of θ and N and is defined as:

$$ARL = E_{\theta}[N] \quad (2)$$

Where it is the expected number of observations that must be recorded before a change-point is incorrectly detected. Therefore, a larger value of ARL is preferable.

Similarly, worst case detection delay is a metric that gives an idea of how slow a model will detect a change-point in a worst case scenario.

If labelled change-points are available for a real world dataset or a synthetic dataset is created, then the ground truth change-point vector, τ^* , is known. For example the *Hausdorff* metric can be used. It measures the furthest temporal distance between a predicted change-point $\hat{\tau}$ and τ^* . It is defined as:

Other standard classifier metrics can also be used for comparing $\hat{\tau}$ and τ^* . This includes the F1-Score that is based on a classifier’s precision and recall:

$$F_1(\hat{\tau}, \tau^*) = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

F1-Score is defined as the harmonic mean of precision and recall. Precision is defined as the ratio of true positives (TP) to the number of true positives (TP) and false positives (FP) and recall is defined as the ratio the number of true positives to the number of true positives plus the number of false negatives. F1-Score is best when $F1 = 1$ (perfect precision and recall) and reaches its worst value at $F1 = 0$. Depending on the context, any other classifier evaluation tools such as the Receiver Operating Characteristics Curve and the Precision Recall Curve may be used as well.

1.4 Related Work

In 2005, Desobry et al. [6] developed an on-line kernel change point detection model based on single class support vector machines (ν -SVMs). The authors train a single class support vector on a past set, $\mathbf{x}_{t,1} = x_{t-m_1}, \dots, x_{t-1}$ of size m_1 and train another single class support vector on a future set $\mathbf{x}_{t,2} = x_t, \dots, x_{t+m_2-1}$ of size m_2 . A ratio is then computed between the two sets that acts as the dissimilarity measure in Hilbert space. If the points are sufficiently dissimilar over some predetermined threshold, η , then a change point is assigned to the time splitting the two sets of data. Desobry argues that a dissimilarity measure between kernel projection of points in a Hilbert space should estimate the *density supports* rather than estimate the probability distributions of each set of points.

In 2007, Harchoui and Cappe [7] approached the off-line change point problem with a fixed number of change points by using kernel change point detection. This

was further extended to an unknown number of change points in 2012 by Arlot et al. [2]. Finally, Garreau and Arlot extended this in line of research kernel change points in the off-line setting of detecting change points. Fundamentally, their method is the kernel version of the following least squares optimization problem:

$$J(\tau, \mathbf{y}) = \frac{1}{n} \sum_{k=1}^K (\tau) \sum (Y_i - \hat{Y}_k)^2 + \beta \text{pen}(\tau) \quad (4)$$

The benefits of this off-line kernel change point detection is that it operates on any kind of data for which a kernel that properly reproduces a Hilbert space can be applied. For example, it can be applied to image data, histogram data, as well as data in \mathbb{R}^P . Garreau shows their KCP procedure outputs an off-line segmentation near optimal with high probability. Lastly, the authors recommend choosing the kernel based on best possible signal to noise ratio that the distribution gives based on Δ^2/M^2 . Therefore, some prior knowledge of the data is necessary for choosing the kernel. Finally, a comparison is made to the kernel two sample hypothesis test that is proposed in [10] that uses the kernel maximum mean discrepancy (MMD) B-test statistic for two-sample testing for rejecting null hypothesis of first equation.

1.5 Our Contributions

Lay out the contributions of this work.

Chapter 2

Chapter 2

2.1 Chapter 2 Section

2.1.1 Chapter 2 Subsection

Subsection Test

Bibliography

- [1] Samaneh Aminikhanghahi and Diane J Cook. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017.
- [2] Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. Kernel change-point detection. *arXiv preprint arXiv:1202.3878*, 6, 2012.
- [3] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial mri: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, 2003.
- [4] E Brodsky and Boris S Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.
- [5] Jie Chen and Arjun K Gupta. *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Springer Science & Business Media, 2011.
- [6] Frédéric Desobry, Manuel Davy, and Christian Doncarli. An online kernel change detection algorithm. *IEEE Trans. Signal Processing*, 53(8-2):2961–2974, 2005.
- [7] Zaid Harchaoui and Olivier Cappé. Retrospective mutiple change-point estimation with kernels. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 768–772. IEEE, 2007.
- [8] Marc Lavielle and Gilles Teyssiere. Adaptive detection of multiple change-points in asset price volatility. In *Long memory in economics*, pages 129–156. Springer, 2007.

- [9] Tze-San Lee. Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice*, 4(4):643–662, 2010.
- [10] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. M-statistic for kernel change-point detection. In *Advances in Neural Information Processing Systems*, pages 3366–3374, 2015.
- [11] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. Online bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1833–1837. IEEE, 2013.
- [12] Yue S Niu, Ning Hao, Heping Zhang, et al. Multiple change-point detection: A selective overview. *Statistical Science*, 31(4):611–623, 2016.
- [13] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.
- [14] Andrey Pepelyshev and Aleksey S Polunchenko. Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570*, 2015.
- [15] Walter Andrew Shewhart. *Economic control of quality of manufactured product*. ASQ Quality Press, 1931.
- [16] M Staudacher, S Telser, A Amann, H Hinterhuber, and M Ritsch-Marte. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):582–596, 2005.
- [17] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC, 2014.
- [18] Charles Truong, Laurent Oudre, and Nicolas Vayatis. A review of change point detection methods. *arXiv preprint arXiv:1801.00718*, 2018.

- [19] Ping Yang, Guy Dumont, and John Mark Ansermino. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on biomedical engineering*, 53(11):2211–2219, 2006.