

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 Related Work	3
1.4 Problem Formulation	4
1.4.1 Performance Measures	4
1.5 Our Contributions	4
2 Chapter 2	5
2.1 Chapter 2 Section	5
2.1.1 Chapter 2 Subsection	5

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. Thus the development and use of statistical control charts was developed for quickly out of control processes [9]. This allowed for the real-time detection of changes in variation that indicated a degradation in quality in the production process. Shewart's method was one of the first formal frameworks to solve the problem of detecting changes in a distribution of observations that arrive one at a time. 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 important area for quickly detecting signal changes in heart rate monitoring [13] [10], epilepsy signal segmentation [7], and multi-modal MRI lesion detection [2] to name a few. It is easy to see why in the context of taking decisions for medical treatment, quickly detecting changes to a patient's health is absolutely necessary for any system to be of practical use, while also balancing the accuracy of these detections. Detecting a false positive or missing a detection could have life-threatening consequences and thus any real-time method must be robust to these types of errors.

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 points in the financial literature are sometimes referred to structural breaks, but for this thesis will use the broader term change points.

Look into references from tarokvsky 2015 book: the articles [52, 358] and references therein. We also argue that quickest change point detection schemes can be effectively applied to the analysis of financial data. In particular, quickest change point detection problems are naturally associated with rapid detection of the appearance of an arbitrage in a market [421].

1.2 Characteristics of the change point problem

A number of surveys of the literature already exist [1], 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. Shewart control charts, CUSUM and autoregressive change-point techniques are all parametric techniques based on Gaussian-family of distributions, also add reference fo AR [8]. For an extensive review, see the monograph by Chen and Gupta [4]. In certain settings, it is not possible to leverage information about the data. Therefore, non-parametric techniques that do not assume particular distributional parameters must be used [3].

The second factor is to deciding whether change-point detection should be applied *offline* or *online*. Some algorithms are off-line or batch algorithms in that they are applied in an ex-post fashion after the dataset has been completely acquired [12]. 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 [11]. 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 a more relevant consideration for off-line change point detection, but 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.

Finally, the last point to address is determining exactly what kinds of statistical changes is an algorithm aiming to detect. Many methods focus solely on detecting changes in the mean of a distribution. Other methods focus solely on detecting changes in the variance of a distribution. Finally, some methods are more general and can detect moment changes past the variance and do not focus on any particular one. Methods like kernel change point detection typically operate in this category.

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 Related Work

In 2005, Desobry et al. [5] 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 spitting the two sets of data.

In 2007, Harchoui and Cappe [?] 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. [?]. 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:

The benefits of this 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. In an application setting, the user would use some training set to calibrate the kernel and the penalty parameters, to then be tested with appropriate accuracy measures on some out of sample data set.

Other related work include a 2015 study by Li et al. [6] that propose the use of M-statistics based on the kernel maximum mean discrepancy (MMD) B-test statistic for two-sample testing.

1.4 Problem Formulation

1.4.1 Performance Measures

Add information about what performance measures will be used to determine the utility of algorithm. E.g. time to detection, F1 measure, etc.

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] 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.

- [9] 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.
- [10] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.
- [11] Walter Andrew Shewhart. *Economic control of quality of manufactured product*. ASQ Quality Press, 1931.
- [12] 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.
- [13] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC, 2014.
- [14] Charles Truong, Laurent Oudre, and Nicolas Vayatis. A review of change point detection methods. *arXiv preprint arXiv:1801.00718*, 2018.
- [15] 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.