Guidance in Using Robotic-Arm Assisted Surgical System for Knee Arthroplasty

- Supportive Surgical Tools with Data Science

David Vichansky

Master of Science

Mathematical Sciences

Utrecht University

May 2023

Guidance in Using Robotic-Arm Assisted Surgical System for Knee Arthroplasty

Supportive Surgical Tools with Data Science

by

David Vichansky

to obtain the degree of Master of Science
at Utrecht University,
to be defended publicly on Thursday June 1st, 2023 at 9:30 AM.

Student number:	6819516
Thesis committee:	
Prof. dr. ir. C. W. Oosterlee	UU, supervisor
N. T. Mücke MSc	UU, supervisor
Dr. I. Kryven	UU, second reader

An electronic version of this thesis is available at https://studenttheses.uu.nl/.
This project was carried out in collaboration with 6Gorilla's and CortoClinics.

Abstract

This thesis project concerns the mathematical analysis of surgical data. In total knee arthoplasty (TKA) surgery a robotic-arm assisted surgery (RAS) system is used to guide the surgeon during the procedure. The use of smart instruments in an operating theatre is still akin to black box thinking. The bigger picture in truth is much more intricate than this. A multifaceted and multidisciplined approach is therefore necessary for a more proactive approach within healthcare. To meet this aim we examine smart tools that can support surgeons in their decision making process. We also investigate previous attempts at using smart analytics in surgery for finding learning curves. Together with reviewing mathematical methods that are useful for the clients task in hand, we apply Bayesian change point detection methods to provide valuable insights for the development of new surgical technologies and devices. By understanding the challenges that surgeons face during the learning process, we can better assess the proficiency of surgeons. The work accumulates with a discussion on which technology is best suited to go hand in hand with data science for postoperative analysis of surgical procedures. The Masters thesis project is completed as part of an internship with a Dutch healthcare software as a service (SaaS) provider.

Contents

List of Figures

List of Tables

Preface

This report is a culmination of several months of research I had the pleasure of developing during my internship. The work presented here would not have been possible without the diligent support of my supervisors, Kees Oosterlee and Nikolaj Mücke, the professional advice of my co-workers, Anton Kuijer and Peter Pilot, and the loving care of those dearest to me.

It became more than simply writing a thesis project for me, but rather a case in point to explore the possibilities of a subject that has very much dominated my life over the past few years - data science in healthcare. At times it challenged me, other times it frustrated me, but on the whole this work has been rewarding and helped me to develop. I will always look back with very fond memories on the four years I have spent working within the healthcare industry, alongside the completion of my master's studies.

David Vichansky
Utrecht
May 2023

Introduction

Most surgical fields are traditionally very analog. The use of robotic surgery in arthroplasty has gained popularity in recent years due to its potential to improve surgical accuracy and precision. While the benefits of robotic surgery in arthroplasty are clear, there is a learning curve associated with this technology that surgeons must overcome to achieve the desired results. In this thesis, we will explore the learning curve associated with using robotic surgery in arthroplasty, and discuss the implications of these developments for patient care and surgical practice.

In cases of end-stage arthritis, surgery can offer a good solution and in many cases a total knee replacement (TKR) is performed. There exist three main categories of surgical robots: passive, semi-active and active robots. In this thesis, we focus on a semi-active Stryker Mako robot in orthopedics. Within the setting of the TKR surgical theatre, the semi-active robot can be seen as a smart-instrument. To make better use of the available smart features that come with this technology, all new data must primarily be understood.

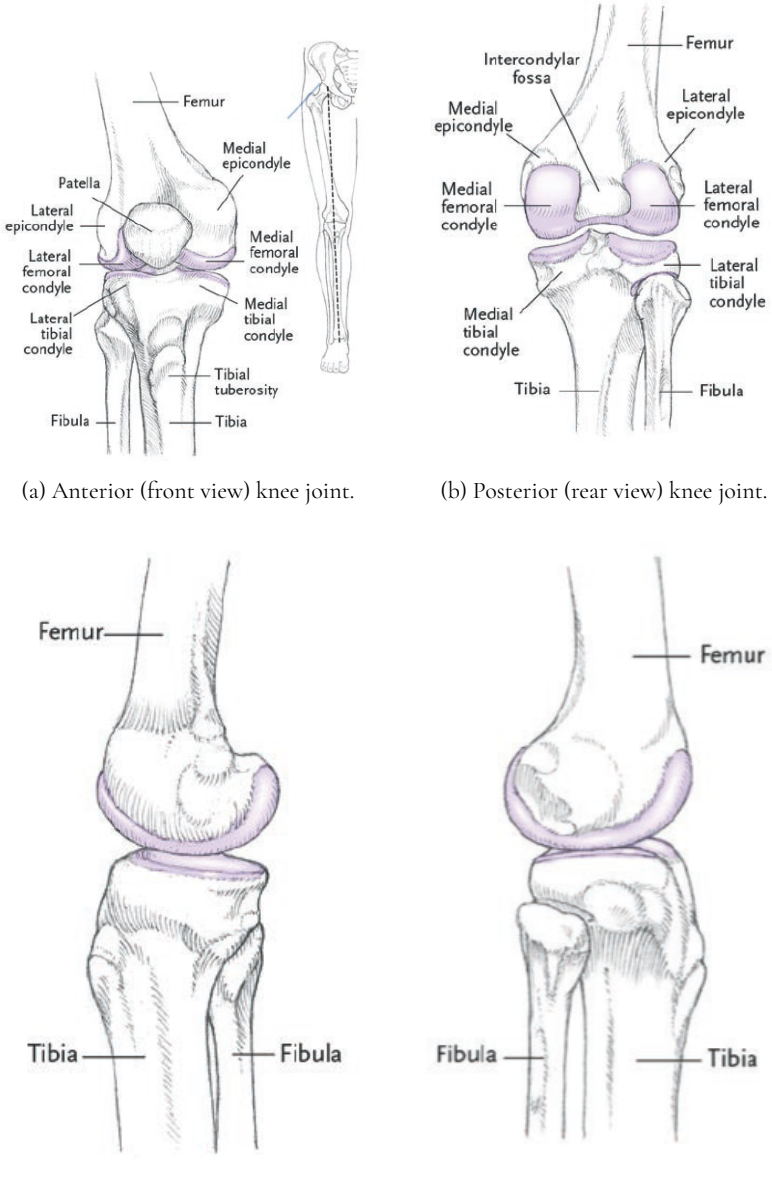
The execution and performance of the surgeon are difficult to measure. With the introduction of surgical robots, digital solutions enter this world. One of the key challenges associated with the learning curve of robotic surgery in arthroplasty is the need for specialised training. Surgeons must undergo training in the use of the robotic system to gain proficiency in its use. This training can be time consuming and costly, requiring the surgeons to take time away from their regular clinical duties. The major skepticism around the introduction of surgical robots is the high costs associated with it and the lack of proof in the added value it provides.

Another challenge associated with the learning curve of robotic surgery in arthroplasty is the need to adapt to the unique features of the robotic system. While traditional arthroplasty procedures rely on the surgeons tactile feedback to guide the surgery, robotic systems rely on visual feedback and computerised control. This shift in approach can be challenging for surgeons who are accustomed to traditional surgical techniques and require time and practice to adapt to the new approach. The knee is also a complex anatomic structure. This results in a lot of compromises having to be made during surgery. To better understand this it is good to start with the anatomy and biomechanics of the knee joint, before proceeding to establish a link between this type of surgery and data science.

Knee replacement surgery

The knee is made up of the femur (thigh bone), tibia (shin bone) and patella (kneecap). The two connecting tissues in the knee are the articular cartilage, the smooth cartilage covering the end of bones which facilitates transmission of loads and prevents the bones from rubbing together for easier movement, and the meniscus, the soft cartilage between the femur and tibia that serves as a cushion and helps absorb shock during motion. Figure 1 illustrates how the knee joint is formed. Gradual mechanical wear damages and thins out the cartilage, resulting in osteoarthritis (OA) of the knee. The result is restricted motion in the joint of the patient. In such aforementioned cases, total knee arthoplasty (TKA) surgery is performed to relieve pain, restore the alignment and function of the knee.

The Mako robotic-arm assisted surgery (RAS) is a technology developed for knee replacement surgery. Computed tomography (CT) scans of the damaged knee are taken to construct a 3D virtual model of



(a) Anterior (front view) knee joint.

(b) Posterior (rear view) knee joint.

(c) Medial (internal view) knee joint.

(d) Lateral (external view) knee joint.

Figure 1: Anatomy of human right knee joint. Image taken from [1]. Cartilage in the knee provides protective cushioning between the femur and tibia bones. In the event of worn cartilage, arthroplasty surgery replaces the damaged area with artificial prosthesis components.

the patients anatomy, enabling the surgeon to create a personalised surgical plan for every patient. The pre-surgical plan will in turn be used intraoperatively to help guide the surgeon in performing the joint replacement procedure, thus allowing for more accurate alignment of the implant. At the start of the surgery the virtual world of the Mako robot has to be connected to the real world, being that of the patient. This is done by placing metal arrays (pins equipped with sensors) into the femur and tibia bones, equipped with haptic technology to communicate with the robotic arm on the precise positioning of the joint.

During surgery the Mako robot arm, operated by the surgeon, holds the surgical instrument. Coupled with limiting saw blade action outside of the haptic boundary, the robotic arm guides the surgeon to cut less (less soft tissue damage and greater bone preservation as compared to manual surgery [2]) for more precise bone removal of the damaged area, in which error in cutting is minimised to under 1 millimetre.

Collaboration of this sort between robots and humans can expand human capabilities. Orthopedic surgeons using the Mako RAS system are fast transforming the field of knee replacement surgery. In his noted book on mankind [3, p. 404], Yuval Noah-Harari depicts how new technology intertwines with human evolution:

> Nearly all of us are bionic these days, since our natural senses and functions are supplemented by devices such as eyeglasses, pacemakers, orthotics, and even computers and mobile phones.

In a way, the Mako robot-arm acts as a sort of extension of the surgeons arm, whilst transforming orthopedist into bionic orthopedic surgeons of the future.

Despite the aforementioned challenges in execution, performance and training, the benefits of robotic surgery in arthroplasty make it a valuable tool for surgeons. With increased precision and accuracy, robotic surgery can help reduce complications and improve patient outcomes. Surgeons currently adopting the Mako system are both able to achieve lower short term pain scores in patients and more precise bone cuts compared with conventional TKA surgery [4, 5, 6]. However, it is essential that surgeons undergo proper training and take the time necessary to become proficient in its use to ensure its safe and effective implementation. For a more extensive comparison of robot assisted and conventional TKA surgery we refer the reader to [7].

Surgical data science

With this technology expected to increase in popularity worldwide [8, 9, 10], there is a growing need of learning how to maximise the effective use of this novel approach. But this is no easy feat to achieve in the medical domain. Obstacles stem from the high volume and velocity with which data is created [11], multiple stakeholders being involved, complexity of applications and nuanced tasks being tackled using limited resources.

Surgical data science aims to improve the quality of interventional healthcare and its value through the capture, organisation, analysis and modelling of data [12]. It encompasses all clinical disciplines in which patient care requires intervention to manipulate anatomical structures with a diagnostic, prognostic or therapeutic goal, such as surgery. By using tools that smoothly integrate into the clinical workflow, there exist a strong potential to complement human cognition. In Figure 2, this idea of technological evolution in surgical processes is outlined. More than ever before, data science is needed to untap the potential of knowledge amidst a torrent of new data sources.

The reality is that daily tasks in industry are fast becoming more and more data intensive, forming a non-reversible trend. Those who gain the most insight from the data will improve performance, whilst those who do not will lag behind. It is unsurprising to see healthcare administrators beginning to incorporate business analytics (BA) to help with decision making needs and to stand out among the industry [13]. Smaller speciality clinics that are not tied to larger healthcare trusts are the most suited to adapt this change. This is because they are more nimble than their large counterparts, with less bureaucratic red tape and more room to be innovative.

The clinic whom we are partnering understands the importance of integrating decision support systems into the patient care pathway. What we are therefore proposing is to answer the question regarding what happens during surgery, followed by why it happens, all in order to facilitate surgeons with targeted feedback. This would in-turn help arthoplasty surgery become more transparent and less of a black box. The word transparency in this setting is meant in a threefold sense of the word: with surgeons to test whether different conclusions are reached elsewhere, with patients in order to manage expectations, and with data scientists to assess which technology is most suited to support the surgical practice.

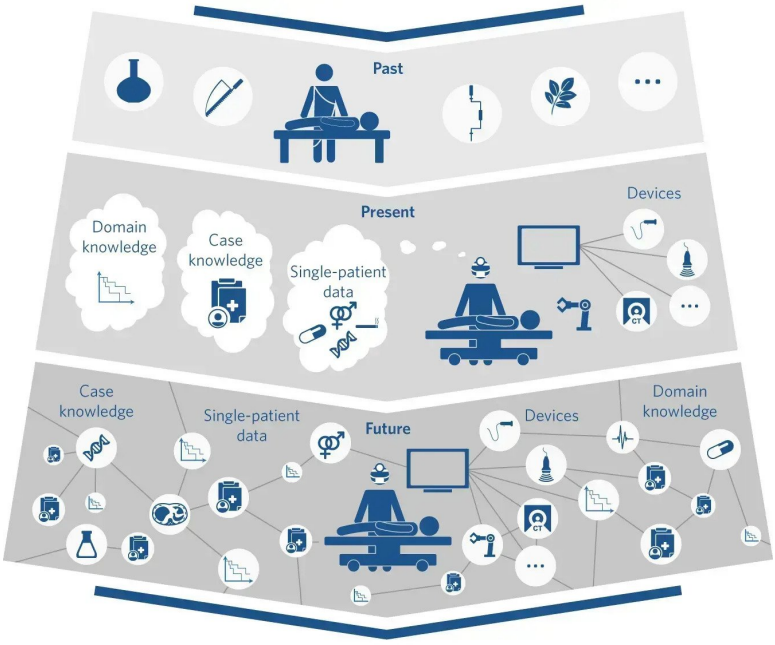


Figure 2: The evolutionary stage of surgical practice has entered into the digital revolution. Image taken from [12]. As more and more digital tools are introduced in surgical settings, the operating surgeon must learn to adapt by becoming more transparent in the digital age.

Marginal gains in surgery

The purpose of small yet significant improvements is to bring monumental changes in results. To accomplish these goals of converting large streams of surgical data into valuable insights, the use of business analytics (BA) is key [11]. Figure 3 depicts the three main characteristics of analytics. This tool is used as a guiding principle to help organisations select the correct analytics capabilities with respect to their operations maturity. With the clinics objective being the enrichment of patient care pathways, progressing surgical analytics up a level is the next phase.

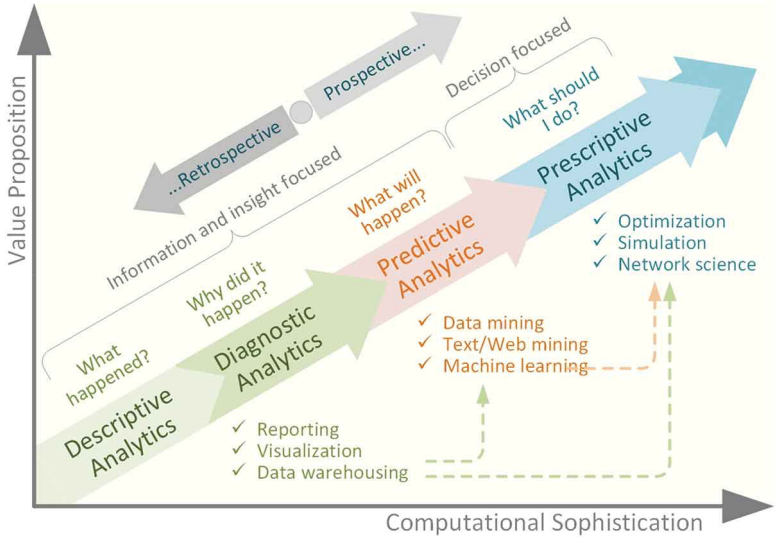


Figure 3: Taxonomy of business analytics (BA) processes leveraged across a hierarchical trajectory. Image taken from [11].

How to marry existing with prospective analytics can be answered by taking a leaf out of a professional cycling book. Sir Dave Brailsford, the mythological manager of Team Sky who dominated professional cycling for the best part of a decade, was one of the instrumental figures in seeing the first Briton being crowned Tour de France champion. A feat that was followed by a further five wins shared among two more British riders.

In many ways, their success is owned largely to the strategy of marginal gains [14, 15]. By looking at the data to see what was happening, to steadily changing their approach and by testing what would happen, the team was finally able to arrive at a decision centric approach and to answer what should be done about this. Evidence of these gains can be witnessed anywhere from using lighter materials in the production of helmets, to personalised nutrition plans for each rider, and even the team cars delivering vital hydration to riders after a long climb but not before so to make the bike lighter and thus the climb easier.

Team Sky were able to capitalise on this success due to the high volume of training and race data being collected and analysed using wearable technology placed on the bikes and riders. Another reason is the ability of the team to look beyond the world of cycling and to technical developments in other fields. An example of such an event was the challenge of solving in-race communication when racing takes place through the mountains. Cross-pollination of ideas eventually took root through the use of military terms as they also communicate in difficult terrains [16].

When looking at the surgical profession, we see many similarities in using data collected with RAS. To draw a parallel, the wearable technology for surgeons is in the form of a Mako robotic arm. This data can reveal, to give an example, the size of an implant required to satisfy correct anatomical alignment of the knee joint. This is akin to riders knowing which strategy to follow depending on which stage in a race they find themselves in. By giving surgeons supportive tools with which to explain to a patient why one strategy is preferred over another, positive change can be brought to the patient care pathway.

In order to build supportive surgical tools with the notion of marginal gains in mind, we must begin work on building a foundation for understanding robotic surgery in arthroplasty. For this reason and with this thesis, inroads in using descriptive and diagnostic analytics are made first. Key challenges associated with this technology are recognised through analysis of the learning curve in robotic surgery, which in turn will help the clinic in producing targeted feedback to surgeons on their progress and areas of improvement.

Chapter 1

Project aim

With the introduction of surgical robots a tsunami of data is brought to the operating theatre. A question that arises is how to deal with the flow of new data? The assumption is that once we start to see which actions are being taken by surgeons and when, we would be better placed to answer the question on what might happen. The focus of this project is on increasing transparency in the understanding between the orthopedics profession and the data science industry. We believe that stronger collaboration between both disciplines can lead to better results. Concurrently, we strive to avoid automation of arthoplasty surgery that can otherwise exasperate a culture of blame. This otherwise threatens the efficacy and safety of implementing predictive technologies in healthcare [17].

In order to continuously improve surgical quality, it is important to have concrete metrics to compare individual surgeon performance. For this reason we perform data science modelling that assists orthopedic surgeons and in turn improves the quality of interventional healthcare. The perioperative data collected during TKA is applied to model the surgical skill curves of surgeons in operating the Mako Stryker RAS system. From there, we extend the learning process to where the various learning phases may lie with the help of data science. Finally, we provide depth to the learning step through analysis of a *golden standard* as a measure for future surgeons in training.

1.1 Problem setup

In this thesis we started a journey to bridge the world of arthoplasty surgery and data scientists. The client is an orthopaedic clinic based in the Netherlands, specialising in treatment of hip and knee osteoarthritis. Owing to their relatively smaller size compared to large hospitals, the clinic thrives by operating efficiently and investing into small improvements that then lead to big results. The clinic was therefore very enthusiastic about the possibility of meaningfully unpacking data from the Mako robot collected during RAS. We help the clinic to analyse surgical performance data produced with the Mako RAS system. This was the start of a longer road with which that clinic aims to bring small improvements to surgeons in using the RAS, that in turn could bring monumental changes in the results of a patient care pathway.

The clinic performs total hip, unicompartmental knee and TKA. We concentrate on the latter because it encompasses over sixty percent of surgeries performed at the clinic. Under the current schedule three surgery days are penciled in per week, with eight surgeries being performed on each of those days. There are three orthopaedic surgeons working in the clinic, all with differing levels of experience. Our findings are based on the most experienced senior surgeon who reaches expert levels in operating the Mako RAS system. This will act as the benchmark category for any future assessment of other surgeons.

In TKA surgery, surgical time can be particularly useful for assessing the proficiency of surgeons performing procedures. Robotic assisted surgery can provide additional metrics and precision during the procedure, making surgical time not the sole factor used to evaluate the success of a knee surgery. Other metrics can

help assess the accuracy of the surgical procedure and the positioning of the implants. These can include but are not limited to: accuracy of implant placement to measure the implants position relative to the patients anatomy, range of motion to assess the degree of movement that the patient has in their knee joint before and after surgery, and ligament balance to measure the tension in the ligaments around the knee joint.

Surgical time is commonly used as a metric for assessing the performance of surgery and surgeons because it is easily measured and provides a simple way to compare the efficiency of different surgical techniques and approaches. Our proposal to the clinic is thus to bridge the understanding between data science and surgery in creating smart decision support tools for the surgeons. We achieve this by applying mathematical tools to data science models for the purposes of bringing transparency and explainability to the surgical time data. Collaboratively, we have come up with four clinical questions that help to match beliefs based on orthopedics intuition. We then assess which data scientific technology is more suited to go hand in hand with TKA surgery through evaluation of surgical times.

1.2 Background and related work

In traditional medicine the operation room is more or less a black box. Conventional analytics within healthcare have tended to focus on randomised control trials (RCT) to measure effectiveness. Other reactive decision support systems may report on the infection rate or a surgical error. The bigger picture in truth is much more intricate than this. A multifaceted and multidisciplined approach is therefore necessary for a more proactive approach within healthcare. Incorporating lessons from aviation [18, 19, 20] and data driven sports [17, 21], such as cycling and Formula 1, can be helpful to better understand the potential offered by RAS.

Success in sports is best characterised by winning a championship, but standardising success in healthcare is not as straightforward [21]. Not only must black box analysis be conducted to understand why a surgical error took place but if we want to also understand which of these features are most prevalent, or influential in other such surgeries, then this is the difficulty we encounter. To aid in our developmental process, we conceptualise the following example: suppose that a reactive model is very good at identifying errors in surgery. The model would notify a member of the clinical team to act upon accordingly in order to fix the error postoperatively. But this would constitute as having lost the race in the Tour de France. Instead, our model should also assist the surgeon in understanding how this mistake can be avoided intraoperatively. This is equivalent to recommending the optimal race strategy to a cyclist in order to maintain a winning position.

How could the explainability part of what is happening during surgical procedures be incorporated inside decision support models to aid surgeons? To answer this question we investigate previous attempts at using smart analytics in surgery, together with reviewing mathematical methods that are useful for our task in hand and how these relate to the work produced for this project.

1.2.1 Knowledge engineering in arthoplasty

The challenges in evaluating surgical skills of trainee surgeons stem from a lack of accurate evaluation metrics, reliable task repetition to verify performance and financial implications of training new surgeons. But at the same time the global market for orthopedic surgical robots expected to grow at 20.75% compounded annual growth rate (CAGR) by 2030 [22], a unique opportunity presents itself whereby surgical variables are recorded using a computer.

Simulation based assessment tools under robot-human collaboration had been extensively studied. Using a purpose built robot, participants haptic skills levels were assessed with Bayesian estimation in laboratory settings [23]. Its findings had shown the greatest decrease in positional distance from the target for low-skilled participants, thus benefiting trainees the most. Knot tying and peg transfer tasks for dexterity assessment of medical students had used assemble at-home surgical box kit. Mobile phone recordings of imitated surgical

sub-tasks had shown training can aid minimally invasive surgery (MIS) skills [24].

Segmentation of surgical tasks using time and motion study was used to evaluate robotic MIS skills [25]. Albeit a simplistic definition of motion systems had been used, stark distinctions could be seen between novice and expert skill levels. Much attention has also been paid to analysing surgical time series data in using the da Vinci surgical system [26]. Multivariate time series had been leveraged to provide novice surgeons with feedback on their progress and improve their knowledge. This work provided a foundation for developing ideas around comparative and objective evaluation tools for the purpose of surgical skill feedback. However, an issue exists around which form this feedback will take on. Recorded surgical videos of experts and novices may be out of sync with each other and comparison between different trials very difficult if not impossible.

In deep learning, convolutional neural network (CNN) model had been trained to classify a participants surgical skill level during knot tying exercises and achieved very high accuracy [27, 28]. An evaluation algorithm to study trajectories of instrument tip in laparoscopic surgery compared similarity between trajectories with dynamic time-warping (DTW) [29]. However, whilst these methods are able to objectively evaluate and provide trainees with skills feedback, the exercises lack real world case validation.

1.2.2 Mathematical application

The clinic wanted to find out how the performance of the house surgeons rivals that of others operating the Mako RAS system. Prior work concentrated with testing of an offline model using classical statistical techniques [30]. This method focused on a single change point inside a learning curve using cumulative summation (CUSUM) analysis that helps detect variation in the trend. CUSUM is a sequential analysis technique that records a running total of deviations between the time series observed values and the target values. An inflection point is observed when a transition between the *inexperienced* and *proficient* phases takes place.

With CUSUM analysis a learning curve had been found for surgical time in Kayani et al. [30], Tay et al. [31] and Vermue et al. [32] but not for either component alignment or ligament balancing. We intend on determining a learning curve for the clinic house surgeons and to assess whether their performance matches the findings in the paper. What makes the learning curves in this thesis unique is the surgeon performed the surgeries before and after the start of the Coronavirus (Covid - 19) pandemic.

As it stands the field of orthopedic surgery is set on its way by using classical statistical techniques in analysing surgeons performance. The CUSUM method is useful for detecting one change point but is unable to generalise to many changepoints or to multivariate data. Furthermore, these findings were based on small data sets. The data we work with is at least threefold larger and can therefore provide more answers in determining whether learning curves exist for surgical tasks.

With the onset of data science many more solutions become available to make use of two such methods involving the use of Bayesian statistics to analyse offline data as presented by Fearnhead [33, 34] and for online data by Adams et al. [35]. Both these methods are handy for segmenting of time series data into multiple segments. It also becomes possible to work with multivariate time series data by updating the BOCD model parameters as shown recently in Wang et al. [36]. This method adapted for surgical tasks can prove useful for assessing the various surgical steps simultaneously. With that we can extend the work further and to build a more complete picture of learning to perform robotic surgery instead of taking each step as a standalone task.

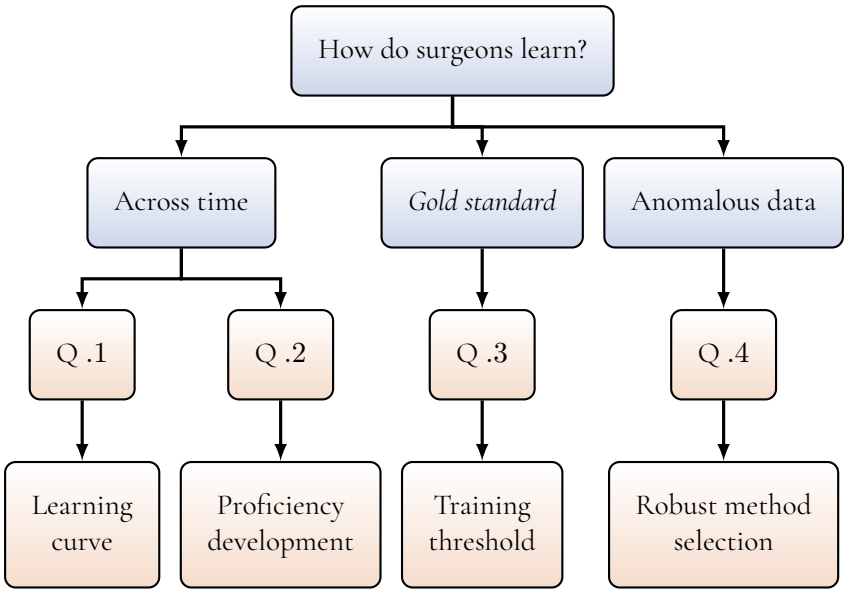
To make the online method work successfully, previous work by Sholihat et al. [37] had also focused on choosing the best hyperparameter settings. However, the questions surrounding whether the offline and online methods are prone to deterioration in their performance when outliers are present is unanswered. Because complications in surgery exist, these reflect anomalous points on time series data. We therefore extend the work of this thesis to analyse the important question on how both methods perform with anomalous data points.

1.3 Studied questions

To help the clinic in beginning to understand and extract meaning from the data it was important to firstly explain which mathematical tools are most suited for which surgical tasks. The purpose of such analysis is to bring efficiency to the clinic. If the knowledge of a learning curve is available, then the expert surgeon can spend less time on being in the operative room training the trainees. Instead, the expert surgeon can perform other tasks, such as screening of potential patients, or assist with patient pathway care.

This thesis therefore inspects past attempts of uncovering learning curves using classical statistics [30, 31, 32]. Whilst the aim being to also grow surgical supportive tools, a comparison using Bayesian inference on streaming data is then made. Ultimately, there is greater benefit of analysing the data as soon as it becomes available in a surgical setting. Once we uncover answers around the issues of how do the surgeons learn, particularly since TKR lacks a true gold standard, we make recommendations to the clinic on how to perform the analysis for all surgical steps in conjunction.

The thesis theme and project questions are summarised as follows:



Question 1: Where is the learning curve?

The clinic wanted us to find the learning curves for the three surgeons performing various surgical tasks. A learning curve is associated with the integration of new techniques. It is therefore important for surgeons who are incorporating these techniques into their surgery to understand what the reported learning curve might mean for them and their patients. After speaking with the surgeons, we understood that the two most difficult skills to hone when training in using the Mako robot are the registration of bone time and ligament balancing time. Therefore, we will test to identify where the learning curve for these two tasks could lie.

In order to uncover the learning curves we will use cumulative summation (CUSUM) analyses to find the flexion points within learning curves. A second comparative method will make use of Bayesian online change point detection [35]. Given the streaming nature of the surgical data, akin to constant inflow of stock market data, we in essence want to assess learning curves on an online model. The comparison of past work done by [30, 31, 32] with offline models will assist in helping to answer the question on which methods suits best to model the surgical skill curves of a surgeon. Our aim is to fit a model that is robust against overfitting.

Question 2: Do surgeons improve over time?

Skill acquisition is the art of improving performance with time because individuals are able to advance their skills from initial learning to improved proficiency. The assumption here is that this also applies to a surgeon who is constantly learning to hone their skills in using the new Mako technology. We will test whether this assumption holds for the time it takes to perform the total surgery, implant planning, ligament balancing, bone registration, total bone cutting and total bone sawing time.

To answer whether surgeons show an improvement in their skill set we opt for comparison between Bayesian offline and online testing. Posterior densities provide us with much more convenient mathematical concepts to grasp than with classical statistics, an advantage of using Bayesian statistics inside of an online algorithm intended for use with streaming data which we aim to leverage.

Question 3: Does a *gold standard* for surgery exist?

The clinic were curious for us to uncover where the *gold standard* may exist. By answering this question we aim to improve the workflow of orthopaedic surgeons in RAS via knowledge utilisation of their performance data. Comparisons can then be made between their recent completed surgery and the gold standard that should help steer the surgeon in the more appropriate direction. We expect the biggest difference to exist during the early training stages but this should not be perceived as negative because it can indicate progress. Due to the complexity and multitude of tasks in surgery, we want to highlight which skills a trainee surgeon has honed and which skills they must still improve on.

Bayesian parameter estimation from the findings of the previous two questions can be leveraged to answer whether a surgeon meets the gold standard. We thus propose to answer this question using univariate Bayesian analysis of surgeons improving across different tasks with time. We then fit a multivariate Bayesian model that assesses overall surgery performance using perioperative and intraoperative recorded surgical steps. The Bayesian approach is most suiting in this setting because the clinic receives streaming data as output from each surgery that requires analysing as a whole to reach a gold standard.

Question 4: How to evaluate surgical performance alongside *outliers*?

The surgeons had explained that one goal of TKA is the balanced tension within the knee throughout range of motions. Depending on the severity of each patients case, the intraoperative surgical procedure time can therefore fluctuate. This can bring distortions to the time series of the overall surgery and on a specific task level. It is therefore important to understand which method is more robust to outliers and is able to discover learning phase correctly. We intend on evaluating both the offline and online methods across a set of experiments to see how each method fares in terms of precision accuracy.

1.4 Data

The surgical data set currently contains 446 patients and encompasses a period of surgeries spanning three years. To help the clinic in beginning to understand and extract meaning from the data, it is important to firstly uncover answers around the issues of how do the surgeons learn. Using time series data is straightforward because it provides detailed understanding of how various aspects of a surgery change over time. In addition, time series data can be used to monitor and improve various aspects of the surgical process itself. For instance, surgeons and healthcare professionals can use time series data to track changes in surgical duration, complications and patient satisfaction over time.

In answering questions $1 - 3$ we will be using the time log data of tasks performed perioperatively. These are recorded in minutes and are reported as real numbers. The six instances of these fields include in the analysis are total surgery time, implant planning time, ligament balancing time, bone registration time, bone saw time and bone cut time. For question 4 we will be using a synthetically constructed data. The

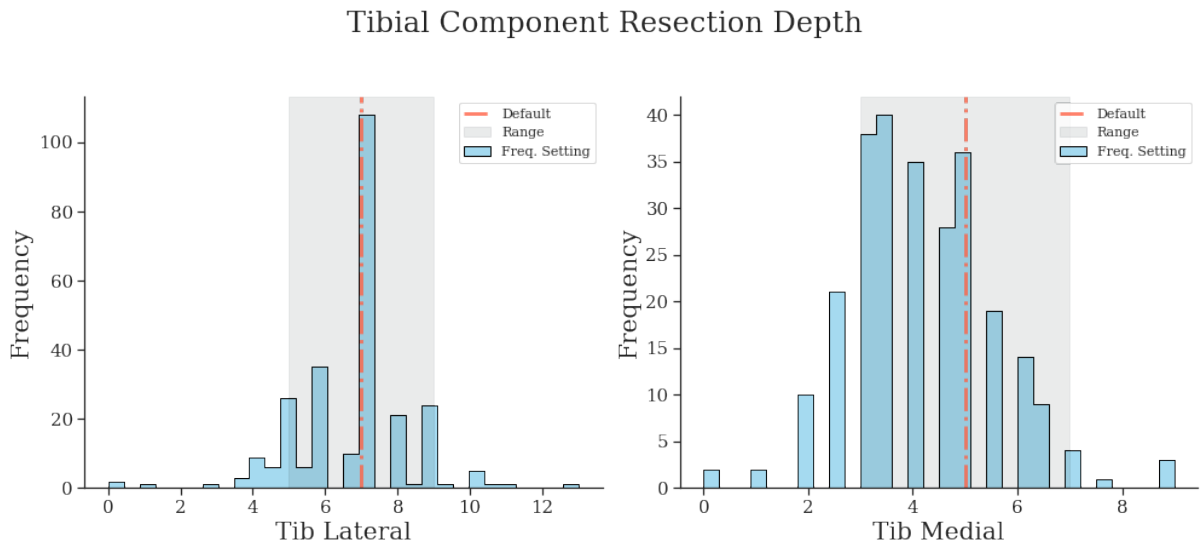


Figure 1.1: Tibial component resection depth for surgeries completed by Surgeon 1 analysed against the recommended Mako RAS settings as found in [38].

characteristics of this data match that of the real world surgical data, only with the ground truth in terms of the location of changepoints and anomalies known to us prior.

While time series data can be a useful tool for analysing surgeries, there are also some potential downsides to consider. Time series data can be complex and difficult to interpret, especially for those without specialised statistical training. This can make it challenging for healthcare professionals to draw meaningful conclusions from the data and make informed decisions about patient care. While time series data can provide insights into how various aspects of a surgery change over time, it may not capture all of the relevant variables that contribute to surgical outcomes. For example, the data may not capture the patients overall health status or the surgeons skill level, which can also play a significant role in surgical outcomes.

A good indication of how a surgeon is performing, whilst also simultaneously allowing for intraprofessional comparison between surgeons in other clinics, is an accuracy metric as shown in Figure 1.1. Here we plot the distribution of bone resection depth values against the recommended setting to use with the Mako RAS.

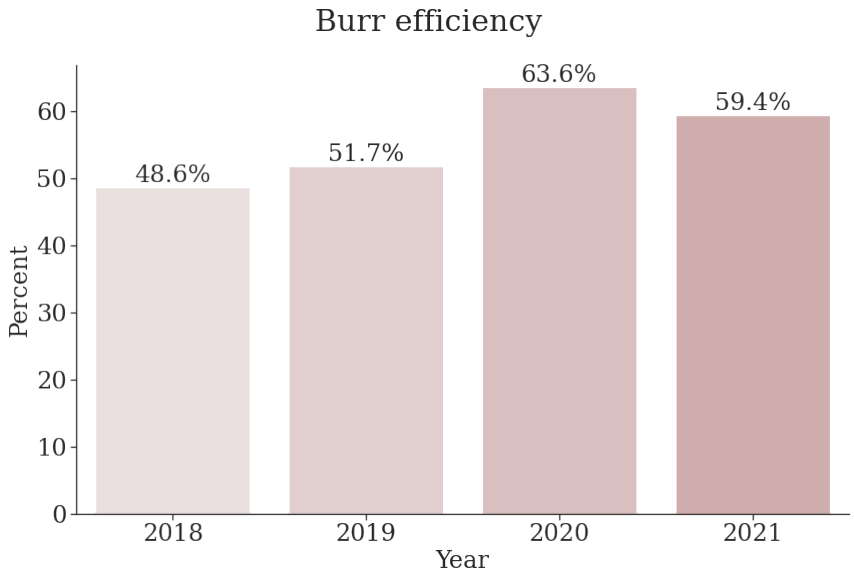


Figure 1.2: Burr time efficiency computed as the bone sawing time out of the bone cutting time.

This however does not show how progress over time. It also fails to take into account patients different anatomical structures or human instinct, which can explain values outside of the prescribed range.

In many contexts, being fast is often associated with being efficient and productive. This is particularly true in the workplace, where employees are often evaluated based on how quickly they can complete tasks or meet deadlines. Similarly in competitive sports, being fast is often associated with being a top performer. Efficiency data can therefore provide a detailed and nuanced understanding of the surgical process. We illustrate in Figure 1.2 the percentage of time the burr blade is actively on. Overall the trend is upward and efficiency is on the rise. But a decrease in efficiency for the year 2021 does not tell us anything about the total time the burr was in operation.

It is important therefore to recognise the limitations of standalone data types and to use it in conjunction with other types of data that require a clinical judgment. This also requires more complex models that are outside the scope of this thesis. Instead, we attempt to incorporate incremental changes in arthoplasty surgery by borrowing from the sporting worlds marginal gains theory. This pushes us to use a universally accepted metric for changes over time.

Chapter 2

Mathematical methods

With this chapter we provide an overview of Bayesian statistical methods used through this project. The work presented here is a summary of the introductory methods which can be found in most introduction to statistics text books. For the purpose of this chapter we extensively relied on *Pattern Recognition and Machine Learning* by Bishop [39].

2.1 Bayesian statistics

Probability theory provides a consistent framework for the quantification and manipulation of uncertainty [39]. The frequentist statistical approach conducts an experiment to infer the probability of an uncertain event A taking place by drawing conclusions based on previous observations. It asks what would happen if an experiment was repeated many times and thus determines the properties of an underlying distribution via the observed data. A frequentist view is thus based on the estimation for probability of an event occurring from a random sample. Conversely, Bayesian statistical approach is based on some limited knowledge which is constantly updated when new observations are incorporated inside the model. It answers what is the probability that an event will take place given the previously observed data. It is considered to be a more robust method as it is less prone to errors [40].

The two fundamental rules of probability theory are the *sum rule* and *product rule*:

$$\mathbb{P}(A) = \sum_{j}^{n} \mathbb{P}(A, C_j) \tag{2.1}$$

$$\mathbb{P}(A, C) = \mathbb{P}(C|A)\mathbb{P}(A) \tag{2.2}$$

We are able to make use of the *marginalisation* technique in Equation 2.1 to sum over every possible parameter value of a random variable A that has a joint distribution with some other random variable C. Probability $\mathbb{P}(A)$ is thus also known as the *marginal probability* because it is obtained by marginalising out other variable C. Equation 2.2 is the *joint probability* of two events occurring as a fraction of all possible outcomes and takes inside of it the *conditional probability* of an event C given A. Owing to the symmetry property $\mathbb{P}(A, C) = \mathbb{P}(C, A)$ we arrive at a solution to a conditional probability using Bayes' theorem.

Theorem 1 (Bayes' rule)**.** *The probability of an event A occurring, given that event C has subsequently occurred, is*

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(C|A)\mathbb{P}(A)}{\mathbb{P}(C)} \tag{2.3}$$

The Bayesian statistical approach on the probability of event A taking place is thus dependent on some belief measure around event C, assuming that this event is known to us. The denominator inside Bayes'

theorem may also be expressed as:

$$\mathbb{P}(C) = \sum_{j}^{n} \mathbb{P}(C|A_j)\mathbb{P}(A_j) \tag{2.4}$$

Or equivalently with the symbol $\neg$ being used to denote an event *not* taking place:

$$\mathbb{P}(C) = \mathbb{P}(C|A)\mathbb{P}(A) + \mathbb{P}(C|\neg A)\mathbb{P}(\neg A) \tag{2.5}$$

It is viewed as a normalising constant which ensures that the sum of the conditional probability over the sum of all values of C is equal to one. Additionally, it is possible to forego the denominator completely by marginalising over a third random variable B as such:

$$\begin{aligned} \mathbb{P}(A|C) &= \frac{\mathbb{P}(A, C)}{\mathbb{P}(C)} \\ &= \frac{\sum_j^n \mathbb{P}(A, B_j, C)}{\mathbb{P}(C)} \\ &= \frac{\sum_j^n \mathbb{P}(A|B_j, C)\mathbb{P}(B_j|C)\mathbb{P}(C)}{\mathbb{P}(C)} \\ &= \sum_{j}^{n} \mathbb{P}(A|B_j, C)\mathbb{P}(B_j|C) \end{aligned} \tag{2.6}$$

This notation will be helpful when we attempt to understand how Bayesian algorithms, in particular BOCD, marginalise over the observed data to predict the next data point.

2.2 Bayesian inference

Using the newly established notions we traverse the landscape to show how to use Bayes' theorem in making Bayesian inference. The Bayesian approach differs from the frequentist method for inference in its use of a *prior distribution* $\mathbb{P}(\theta)$ to express the uncertainty present before seeing the data [41]. We therefore do not assume the model parameters θ to be fixed and are interested in quantifying the uncertainty around these parameters. This is done via specifying the probability at each value that the parameters can take up. The goal of a Bayesian inference process is to then allow the uncertainty remaining after observing the data $\mathbf{x}$ to be expressed in the derivation of the *posterior distribution*:

$$\mathbb{P}(\theta|\mathbf{x}) = \frac{\mathbb{P}(\mathbf{x}|\theta)\mathbb{P}(\theta)}{\mathbb{P}(\mathbf{x})} \tag{2.7}$$

This method clearly incorporates previously attained knowledge of the parameter values. We can forego the normalising constant in Equation 2.7 and write using proportionality:

$$\mathbb{P}(\theta|\mathbf{x}) \propto \mathbb{P}(\mathbf{x}|\theta)\mathbb{P}(\theta) \tag{2.8}$$

The posterior is then simply represented as the probability likelihood that under some model parameters θ we would observe the data $\mathbf{x}$, multiplied with the prior containing some preexisting knowledge about the parameter values.

2.2.1 Likelihood function

The interpretation for the likelihood in Bayesian inference is to uncover the underlying probability model (UPM) that gives the highest probability of seeing the data. It borrows itself from the frequentist approach

of maximum likelihood estimation (MLE) for estimating some variable. In the frequentist setting, the MLE of the parameter we want to infer:

$$\begin{aligned}\theta_{\text{MLE}} &= \arg\max_{\theta} \mathbb{P}(\mathbf{x}|\theta) \\ &= \arg\max_{\theta} \prod_{i} \mathbb{P}(x_i|\theta)\end{aligned} \tag{2.9}$$

Taking the product of infinitely many probabilities rapidly brings the MLE towards zero. This is why we prefer to maximise the log function instead:

$$\begin{aligned}\theta_{\text{MLE}} &= \arg\max_{\theta} \log \mathbb{P}(\mathbf{x}|\theta) \\ &= \arg\max_{\theta} \log \prod_{i} \mathbb{P}(x_i|\theta) \\ &= \arg\max_{\theta} \sum_{i} \log \mathbb{P}(x_i|\theta)\end{aligned} \tag{2.10}$$

To solve for this suppose then that our data is drawn from a Gaussian distribution with parameters $\theta = [\mu, \sigma]$. The likelihood term is:

$$\mathbb{P}(\mathbf{x}|\mu, \sigma) = \prod_{i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \tag{2.11}$$

Maximising the log of the likelihood function:

$$\begin{aligned}\log \mathbb{P}(\mathbf{x}|\mu, \sigma) &= \log\left(\prod_{i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)\right) \\ &= \log\left(\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2\right)\right) \\ &= \log\left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\right) + \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2\right)\right) \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2 \\ &= -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2\end{aligned} \tag{2.12}$$

Taking the derivative with respect to model parameters to find the MLE of the mean:

$$\begin{aligned}\frac{\partial}{\partial\mu}\log \mathbb{P}(\mathbf{x}|\mu, \sigma) &= \frac{\partial}{\partial\mu}\left(-\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2\right) \\ 0 &= \frac{1}{\sigma^2}\sum_{i}(x_i - \mu) \\ \hat{\mu}_{\text{ML}} &= \frac{1}{n}\sum_{i} x_i\end{aligned} \tag{2.13}$$

Because by equating the first order derivative to zero one can find the extremum point. Similarly the MLE of the variance:

$$\frac{\partial}{\partial\sigma^2}\log \mathbb{P}(\mathbf{x}|\mu, \sigma) = \frac{\partial}{\partial\sigma^2}\left(-\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i}(x_i - \mu)^2\right) \tag{2.14}$$

$$0 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (x_i - \mu)^2$$

$$\hat{\sigma}^2_{\text{ML}} = \frac{1}{n}\sum_i (x_i - \mu)^2$$

Thus we obtain the sample mean mean $\hat{\mu}_{\text{ML}}$ and sample variance $\hat{\sigma}^2_{\text{ML}}$.

2.2.2 Maximum a posteriori probability

Now we are ready to draw parallels between the MLE and maximum a priori (MAP) estimation. This method selects the most likely set of parameters for the posterior distribution $\mathbb{P}(\theta|\mathbf{x})$ conditional on the data. MAP is preferred under scenarios where we possess some prior knowledge $\mathbb{P}(\theta)$ and want to incorporate this into the model by weighing the likelihood function. We also assume here that the variance parameter σ^2 is known and try to estimate via maximising the unknown population mean μ. It is a reasonable assumption to make that our mean is from a univariate Normal distribution because it is often the case that our variance is within some range but the mean is unknown.

We proceed by taking Equation 2.8 and want to solve for the MAP of the posterior distribution:

$$\mathbb{P}(\mu|\mathbf{x}) \propto \mathbb{P}(\mathbf{x}|\mu)\mathbb{P}(\mu) \tag{2.15}$$

Note how once again because the denominator $\mathbb{P}(\mathbf{x})$ does not depend on the parameter μ, we omit it and instead maximise the numerator. Observing the likelihood function is simply Equation 2.11 and consulting Bishop [39, p.97-98], we see that the likelihood function takes the form of the exponential of a quadratic form in μ. If we multiply this likelihood by another Gaussian process, we will obtain another Gaussian. This is true since each Gaussian can be written as an exponential multiplied with a quadratic, the product of which is another exponential of a quadratic form. We therefore take the *natural conjugate prior* to have the same form as the likelihood:

$$\mathbb{P}(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \tag{2.16}$$

Where μ_0, σ_0^2 are parameters of the prior distribution. For detailed explanation and derivation see Appendix A. Here we show how to derive the MAP estimator for the parameter mean μ. Multiplying the likelihood with the prior to obtain the posterior:

$$\mathbb{P}(\mu|\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2\right) \cdot \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \tag{2.17}$$

We can then take the logarithm:

$$\log \mathbb{P}(\mu|\mathbf{x}) = -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2 - \log\left(\sqrt{2\pi\sigma_0^2}\right) - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 \tag{2.18}$$

Then differentiate with respect to parameter μ which we want to maximise:

$$\frac{\partial}{\partial\mu}\log \mathbb{P}(\mu|\mathbf{x}) = \frac{\partial}{\partial\mu}\left(-n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2 - \log\left(\sqrt{2\pi\sigma_0^2}\right) - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \tag{2.19}$$

Setting this equal to zero and solving:

$$0 = \frac{1}{\sigma^2}\sum_i (x_i - \mu) - \frac{1}{\sigma_0^2}(\mu - \mu_0) \tag{2.20}$$

$$\frac{1}{\sigma_0^2}\mu - \frac{1}{\sigma_0^2}\mu_0 = \frac{1}{\sigma^2}\sum_i x_i - \frac{1}{\sigma^2}n\mu$$
$$\frac{1}{\sigma_0^2\mu} + \frac{1}{\sigma^2}n\mu = \frac{1}{\sigma^2}\sum_i x_i + \frac{1}{\sigma_0^2}\mu_0$$
$$\frac{1}{\sigma^2\sigma_0^2}\mu\left(\sigma^2 + n\sigma_0^2\right) = \frac{1}{\sigma^2\sigma_0^2}\left(\sigma_0^2\sum_i x_i + \sigma^2\mu_0\right)$$

The MAP estimator is therefore solved with:

$$\mu_{\text{MAP}} = \frac{\sigma^2\mu_0 + n\sigma_0^2\hat{\mu}_{\text{ML}}}{\sigma^2 + n\sigma_0^2} \tag{2.21}$$
$$\mu_{\text{MAP}} = \sigma_{\text{MAP}}^2\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\hat{\mu}_{\text{ML}}}{\sigma^2}\right)$$

Where σ_{MAP}^2 equals:

$$\sigma_{\text{MAP}}^2 = \frac{1}{\frac{1}{\sigma_0^2+\sigma^2}} \tag{2.22}$$

It is important to point out that with no observed data points, Equation 2.21 reduces to the posterior mean μ_0. But as we gather more data so that $n \to \infty$, the variance σ_{MAP}^2 goes to zero and the MAP estimator becomes the MLE $\hat{\mu}_{\text{ML}}$:

$$\begin{aligned}\mu_{\text{MAP}} &= \frac{\sigma^2\mu_0}{\sigma^2 + n\sigma_0^2} + \frac{n\sigma_0^2\hat{\mu}_{\text{ML}}}{\sigma^2 + n\sigma_0^2} \\ &= \frac{\sigma^2\mu_0}{\sigma^2 + n\sigma_0^2} + \frac{\hat{\mu}_{\text{ML}}}{1 + \frac{\sigma^2}{n\sigma_0^2}} \\ &= \frac{\sigma^2\mu_0}{\sigma^2 + n\sigma_0^2} + \frac{\hat{\mu}_{\text{ML}}}{1 + \frac{\sigma^2}{n\sigma_0^2}} \\ &\quad \to 0 \qquad\qquad \to 0\end{aligned} \tag{2.23}$$

2.2.3 Prior predictive distribution

The focus so far has been on inferring unknown model parameters θ. When we address the *prior predictive*, we are in the business of predicting future observable data values. The *prior predictive distribution* is used to predict future observation x^{new}, given some predefined hyperparameters α, without seeing the data firsthand and by marginalising over the conditional probabilities:

$$\mathbb{P}(x^{\text{new}}|\alpha) = \int \mathbb{P}(x^{\text{new}}|\theta)\mathbb{P}(\theta|\alpha)\ \mathrm{d}\theta \tag{2.24}$$

It is therefore used in assessing how well the prior distribution $\mathbb{P}(\theta|\alpha)$ captures our beliefs which are embedded inside hyperparameters α.

2.2.4 Posterior predictive distribution

The *posterior predictive distribution* follows the sum rule by marginalising over the model parameters θ and leaves us with simply the predictive density of a new data point x^{new} given the observed data $\mathbf{x}$:

$$\mathbb{P}(x^{\text{new}}|\mathbf{x}) = \int \mathbb{P}(x^{\text{new}}|\theta)\mathbb{P}(\theta|\mathbf{x})\ \mathrm{d}\theta \tag{2.25}$$

We also want to avoid performing painstaking integration in Equation 2.25. The idea is to instead leverage conjugacy and pick a conjugate prior with the same functional form as the posterior:

$$\begin{aligned}\mathbb{P}(x^{\text{new}}|\mathbf{x},\alpha) &= \int \mathbb{P}(x^{\text{new}}|\theta)\mathbb{P}(\theta|\mathbf{x},\alpha)\ \mathrm{d}\theta \\ &= \int \mathbb{P}(x^{\text{new}}|\theta)\mathbb{P}(\theta|\beta)\ \mathrm{d}\theta \\ &= \mathbb{P}(x^{\text{new}}|\beta)\end{aligned} \tag{2.26}$$

For some new hyperparameter β. This allows us to deduce the posterior predictive conjugate model without integration.

2.2.5 The exponential family

The business of Bayesian inference is in correctly and efficiently deducing properties regarding a probability distribution by computing the posterior distribution in Equation 2.8. Fortunately, distributions belonging to the *exponential family* (EF) possess many useful qualities that help achieve exactly this. To model the probability distribution $\mathbb{P}(\mathbf{x})$ of an n variables $\mathbf{x} = \{x_1, ..., x_n\}$, the EF are written in the following form:

$$\mathbb{P}(\mathbf{x}|\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta})\exp\left[\boldsymbol{\eta}^{\mathsf{T}}\mathbf{u}(\mathbf{x})\right] \tag{2.27}$$

The non-negative function $h(\mathbf{x})$ by construction determines the support on $\mathbf{x}$ and therefore does not depend on natural parameters $\boldsymbol{\eta}$. The function $g(\boldsymbol{\eta})$ is a normalising function for the distribution and the *sufficient statistic* $\mathbf{u}(\mathbf{x})$ is a function that depends only on $\mathbf{x}$. The EF set of parametric distributions therefore incorporates Gaussian, Poisson and Binomial distributions, just to name a few. But does not include, for instance, the Uniform distribution.

Members of the EF allow for convenient computation of the expectation and variance for a distribution. The identical form of our EF also eliminates the need for us to pick one distribution over another and to watch our model crumble into impracticality due to skewness of the data or values being continuous when we needed them to be discrete. As we shall see later, the use of EF is beneficial to us because it incorporates the sufficient statistic. Thus allowing to fit the model by maximising for the parameters as all the information is contained only in $\mathbf{u}(\mathbf{x})$. Hence there is no need to store all the data, but rather only the values contained inside the sufficient statistic. EF distributions also make use of the *conjugate prior* which makes our lives simpler when iteratively updating and computing for the closed form posterior predictive distribution.

Bishop [39, p.113-117] provides several example use cases on EF distributions. We wish to instead show how to cast the Normal distribution with known variance σ_0^2 into the EF form because it will be more relevant to our case on hand.

Example 1. Let x be a random variable such that $x \sim \mathcal{N}(\mu, \sigma_0^2)$. Then the probability model belonging to an EF is written:

$$\mathbb{P}(x|\boldsymbol{\eta}(\theta)) = h(x)g(\boldsymbol{\eta}(\theta))\exp\left[\boldsymbol{\eta}(\theta)^{\mathsf{T}} \cdot \mathbf{u}(x)\right] \tag{2.28}$$

Note however that only one parameter is unknown, hence $\boldsymbol{\eta}$ ceases being a vector valued function:

$$\mathbb{P}(x|\eta) = h(x) \cdot g(\eta)\exp\left[\eta \cdot \mathbf{u}(x)\right] \tag{2.29}$$

Now taking the standard form of a Normal distribution we bring it to the EF form:

$$\begin{aligned}\mathbb{P}(x|\eta) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot \exp\left[-\frac{1}{2\sigma_0^2}(x-\mu)^2\right] \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot \exp\left[-\frac{1}{2\sigma_0^2}x^2 + \frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2}\right]\end{aligned} \tag{2.30}$$

Because the left term is not dependent on the random variable x we take it outside:

$$\mathbb{P}(x|\eta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot \exp\left[-\frac{1}{2\sigma_0^2}x^2\right] \cdot \exp\left[\frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2}\right] \tag{2.31}$$

By taking the log-normaliser relation $g(\eta) = \exp[-A(\eta)]$ we arrive at an equivalent form for the EF:

$$\mathbb{P}(x|\eta) = h(y) \cdot \exp[\eta \cdot \mathbf{u}(x) - A(\eta)] \tag{2.32}$$

We can immediately see that terms which do not depend on μ go inside $h(x)$ and $\mathbf{u}(x)$ is a sufficient statistic for η. Bunching these terms into their respective EF counterparts we obtain:

$$\begin{aligned} h(x) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot \exp\left[-\frac{1}{2\sigma_0^2}x^2\right] \\ \mathbf{u}(x) &= x \\ \eta(\mu) &= \frac{\mu}{\sigma_0^2} \\ A(\eta) &= \frac{\mu^2}{2\sigma_0^2} \end{aligned} \tag{2.33}$$

Or equivalently:

$$g(\eta) = \exp[-A(\eta)] = \exp\left[-\frac{\mu^2}{2\sigma_0^2}\right] \tag{2.34}$$

At this point we also wish to explain why the natural parameter $\boldsymbol{\eta}(\theta)$ is named as such. This parameter refers to the set of all $\boldsymbol{\eta}(\theta)$ which belong to the natural parameter space. Then from Example 1 above and observing that $\eta(\mu) = \frac{\mu}{\sigma_0^2}$ can take any real value, the natural parameter space is simply $(-\infty, \infty)$.

Sufficient statistics

A sufficient statistic of a sample summarises all the required information about some parameter. This means that no matter how much more data we throw at the sufficient statistic, we gain no more information to do with the unknown model parameters θ. Fisher first introduced the notion of *sufficiency* [42] that dealt with the matter of providing a precise form of a distribution from a randomly drawn sample. The independence between $\mathbf{u}(\mathbf{x})$ and θ was proven by Pitman soon after [43].

In Bayesian statistics we can characterise this notion of independence as follows:

$$\mathbb{P}(\theta|\mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbb{P}(\theta|\mathbf{u}(\mathbf{x})) \tag{2.35}$$

This characterises to us what is essential in the data and what we can disregard. We can therefore extend Example 1 to show how the sufficient statistic works in practice. The benefit of EF, as we will see, is that the MLE depends on the data only through $\sum_i \mathbf{u}(x_i)$ [39, p.117]. It is no accident therefore that this is the sufficient statistic for the Normal distribution.

Example 2. Considering a set of data $\mathbf{x} = \{x_1, ..., x_n\}$ where $x_i \sim \mathcal{N}(\mu, \sigma_0^2)$, the likelihood function is:

$$\mathbb{P}(\mathbf{x}|\boldsymbol{\eta}) = \prod_i h(x_i) \cdot g(\boldsymbol{\eta})^n \exp\left[\boldsymbol{\eta}^\intercal \sum_i \mathbf{u}(x_i)\right] \tag{2.36}$$

Where we have simply rewritten the vector parametrisation form. Then taking the logarithm:

$$\log \mathbb{P}(\mathbf{x}|\eta) = \sum_i \log\left(h(x_i)\right) + n \cdot \log\left(g(\eta)\right) + \eta \cdot \sum_i u(x_i) \tag{2.37}$$

Recall that we wish to differentiate with respect to η in order to maximise this value as we had done earlier:

$$\frac{\partial}{\partial \eta} \log \mathbb{P}(\mathbf{x}|\eta) = \frac{\partial}{\partial \eta}\left(\sum_i \log\left(h(x_i)\right) + n \cdot \log\left(g(\eta)\right) + \eta \cdot \sum_i u(x_i)\right) \tag{2.38}$$

Setting this equal to zero and solving:

$$\begin{aligned} 0 &= n \cdot \log\left(g(\eta_{\text{ML}})\right) + \sum_i u(x_i) \\ -\log\left(g(\eta_{\text{ML}})\right) &= \frac{1}{n}\sum_i u(x_i) \\ A(\eta_{\text{ML}}) &= \frac{1}{n}\sum_i u(x_i) \end{aligned} \tag{2.39}$$

This result is useful to us because we see that the MLE is the optimal estimator since the natural parameter η_{ML} only depends on the data contained inside the set of sufficient statistics $\sum_i u(x_i)$. Once again we are able to store only the data that is required given a finite number of sufficient statistics.

Hence sufficiency allows for $\mathbf{u}(\mathbf{x})$ to be a sufficient statistic for θ since there is no more information regarding θ beyond what is already expressed in $\mathbf{u}(\mathbf{x})$.

Conjugate priors

Under Bayesian inference the Bayes' Rule from Theorem 1 is adapted to compute the posterior distribution of the parameters θ conditional on the data $\mathbf{x}$. A prior is termed a conjugate prior if it is from the same distribution family as the posterior for some corresponding likelihood. This means the posterior distribution has the same functional form as the prior and hands us the advantage of obtaining the MAP function using a made simple derivation. This is exactly how we obtained a closed form solution for the posterior parameters in Equation 2.22 - 2.23.

In the event that no closed form solution exists, conjugate priors allow for sequential learning on the posterior. This allows us to skip for each time step t the computationally expensive multiplication of the likelihood function with the prior inside of Equation 2.25. Instead we can update the parameters to model the posterior probability distribution. What more, once we have calculated posterior for some time step t, it directly becomes the prior at the next time step $t+1$. In summary, conjugate priors sequentially lead us to the best possible parameters θ that maximise the posterior and allow its estimated distribution to approach its true value. Recall that we derived the MAP estimator in Equation 2.17. All that remains is to show with Figure 2.1 how the posterior distribution is updated at each time step t.

Notice how the posterior distribution squeezes around the true value of the mean. The y - axis is being stretched as a result of the distribution peak growing taller because we allow the model to be led by the data at each time step. As more and more data is fed into the sequential updating of our model, the confidence

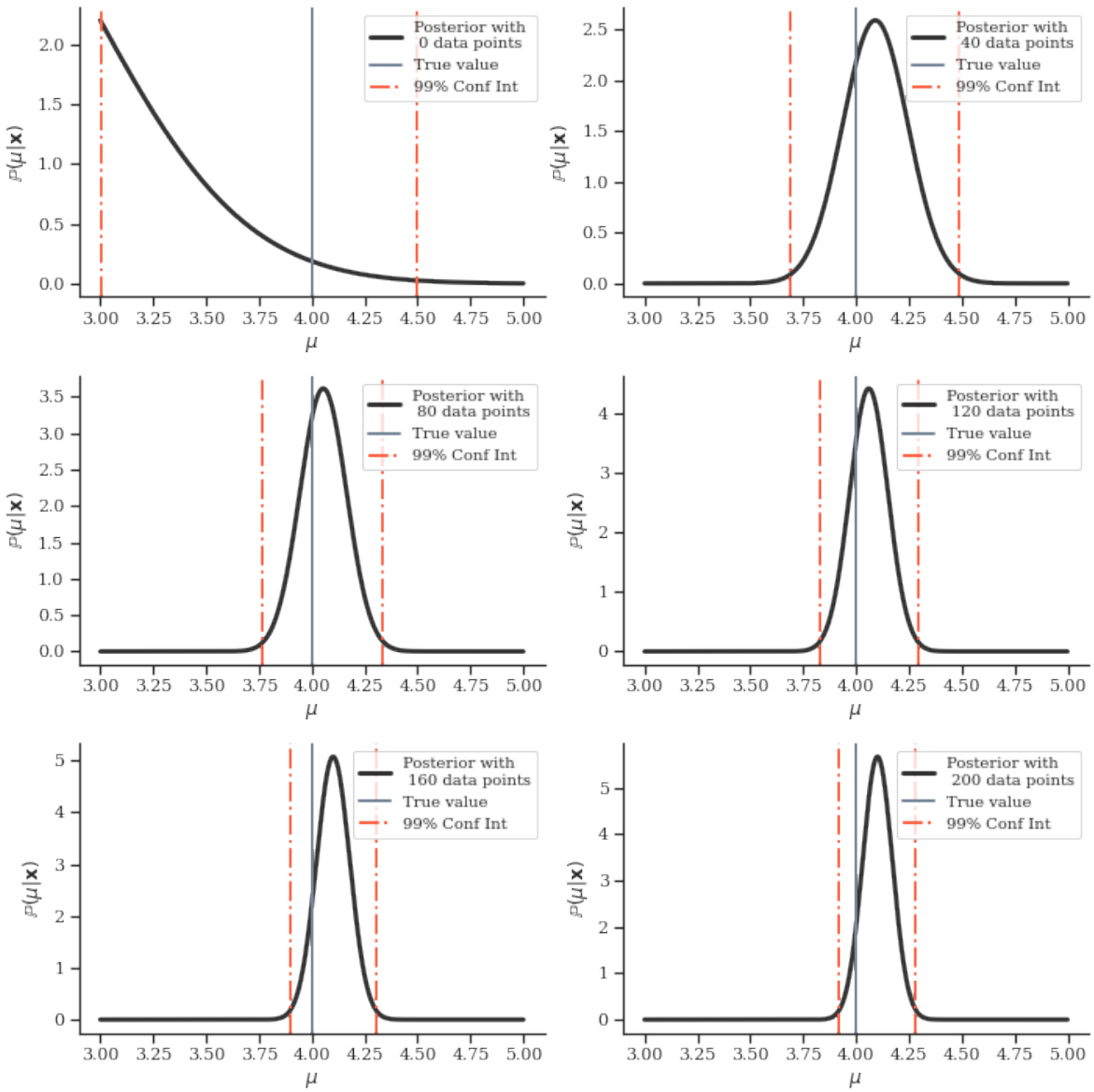


Figure 2.1: Bayesian estimation of the mean of a Gaussian distribution. In this example we use a *strong prior* with $\mathbb{P}(\mu) \sim \mathcal{N}(4, 1)$ to approximate for the posterior distribution mean, depending on some seen data $\mathbf{x}$. At each time step t the prior is multiplied with the likelihood function to obtain the posterior distribution at time t. The posterior is then sequentially fed into an updated prior at time $t + 1$. We then repeat the process of multiplication with the likelihood to obtain posterior at $t + 1$, and so on.

interval also narrows. This provides us with an easy way of interpreting how the parameter values of the prior change with Bayesian updating. Because we now understand what conjugacy provides us with, we can write a conjugate prior of an EF to be of the following form [39, p.117]:

$$\mathbb{P}\left(\boldsymbol{\eta}|\boldsymbol{\chi}, \nu\right) = f\left(\boldsymbol{\chi}, \nu\right) \cdot g(\boldsymbol{\eta})^{\nu} \exp\left[\nu \boldsymbol{\eta}^{\mathsf{T}} \cdot \boldsymbol{\chi}\right] \tag{2.40}$$

Where $\boldsymbol{\chi}, \nu$ are hyperparameters belonging to an EF form.

Then we can multiply the conjugate prior $\mathbb{P}(\boldsymbol{\eta}|\boldsymbol{\chi}, \nu)$ with the likelihood $\mathbb{P}(\mathbf{x}|\boldsymbol{\eta})$ in order to obtain and verify that the posterior distribution has the same functional form as the conjugate prior:

$$\mathbb{P}(\boldsymbol{\eta}|\mathbf{x}, \boldsymbol{\chi}, \eta) = \mathbb{P}(\mathbf{x}|\boldsymbol{\eta}) \cdot \mathbb{P}(\boldsymbol{\eta}|\boldsymbol{\chi}, \nu) \tag{2.41}$$

$$= \left(\prod_i h(x_i) \cdot g(\boldsymbol{\eta})^n \exp \left[\boldsymbol{\eta}^\mathsf{T} \sum_i \mathbf{u}(x_i) \right] \right) \cdot \left(f(\boldsymbol{\chi}, \nu) \cdot g(\boldsymbol{\eta})^\nu \exp \left[\nu \boldsymbol{\eta}^\mathsf{T} \cdot \boldsymbol{\chi} \right] \right)$$
$$= \prod_i h(x_i) \cdot f(\boldsymbol{\chi}, \nu) \cdot g(\boldsymbol{\eta})^{n+\nu} \exp \left[\boldsymbol{\eta}^\mathsf{T} \sum_i \mathbf{u}(x_i) + \boldsymbol{\eta}^\mathsf{T} \nu \boldsymbol{\chi} \right]$$

Because the first two terms are constant with respect to $\boldsymbol{\eta}$, the posterior is proportional to the following:

$$\mathbb{P}(\boldsymbol{\eta} | \mathbf{x}, \boldsymbol{\chi}, \eta) \propto g(\boldsymbol{\eta})^{n+\nu} \exp \left[\boldsymbol{\eta}^\mathsf{T} \sum_i \mathbf{u}(x_i) + \boldsymbol{\eta}^\mathsf{T} \nu \boldsymbol{\chi} \right] \tag{2.42}$$

It therefore holds that the posterior has the same functional form as the prior, just as we had set out to do in the beginning with Equation 2.8, only instead with parameters:

$$\begin{aligned} \nu_{\text{posterior}} &= n + \nu_{\text{prior}} \\ \boldsymbol{\chi}_{\text{posterior}} &= \sum_i \mathbf{u}(x_i) + \nu_{\text{prior}} \cdot \boldsymbol{\chi}_{\text{prior}} \end{aligned} \tag{2.43}$$

Where we interpret ν_{prior} as the number of observations inside the prior which have the value of the sufficient statistic $\mathbf{u}(\mathbf{x})$ given by $\boldsymbol{\chi}$. Finally, to be able to arrive sequentially at some optimal parameters which maximise the posterior as shown in Figure 2.1 all that we do is update the hyperparameters at each time step t as follows:

$$\begin{aligned} \nu_t &= \begin{cases} \nu_{\text{prior}} & \text{if } t = 0 \\ \nu_{t-1} + 1 & \text{if } t > 0 \end{cases} \\ \boldsymbol{\chi}_t &= \begin{cases} \nu_{\text{prior}} \cdot \boldsymbol{\chi}_{\text{prior}} & \text{if } t = 0 \\ \boldsymbol{\chi}_{t-1} + \mathbf{u}(\mathbf{x}_{t-1}) & \text{if } t > 0 \end{cases} \end{aligned} \tag{2.44}$$

2.3 Summary of recursive Bayesian estimation

After discussing in length the many benefits of EF models, the key take away points for use with Bayesian inference and its recursive reconciliation of new data are summarised as follows:

1. EF models allow for inference with a finite number of sufficient statistics, as was shown using $\sum_i u(x_i)$.
2. EF models allow for incremental calculation as new data arrives.
3. Conjugate EF representation allows for EF distribution over $\boldsymbol{\eta}$ that can be summarised with hyperparameters belonging to the same distribution.
4. Inferring the parameter vector $\boldsymbol{\eta}$ associated with the run length r_t is then made simpler using the methods discussed in Section 2.2.

Chapter 3

Change point detection

In this chapter we discuss the CUSUM algorithm used for the analysis of trend detection in learning curves. As a contrasting methodology to the one currently used in industry, we introduce two change point detection algorithms called offline BCD [33] and BOCD [35]. Interested readers should refer to the original papers whilst the discussion points presented here are merely used in explaining reasoning behind the methods used by the authors. We do this by making the connection between the methodology for change point detection and the partitioning of data sequences by computing their posterior distribution over run lengths, which is helped by using the Bayesian methods outlined in Chapter 2.

Change detection for time series data deals with detecting whether a change point has occurred or not. Transition between phases over time of the underlying process can be detected, such as using mean or variance shifts, that can in turn be related to the start of a new phase. Before applying to surgical data, with this chapter we explore the capabilities and limitations of the three aforementioned methods. In many industries offline BCD and BOCD are used in order to alert, for instance, fund managers when sudden or relevant movements occur in the market. We extend the work on CUSUM learning curves through further examination of multivariate data also using both Bayesian methods.

With the RAS system data belonging to a streaming nature reminiscent of an online model, our intuition tells us that there may exist more than one learning curve belonging to each separate partition of the data. By identifying changes in underlying surgical tasks, supervising surgical skills progression can be performed in real time. We are interested in being able to detect the time at which a new phase start because it may be indicative of a surgeon becoming more proficient in a RAS task. Furthermore, we inspect whether the streaming data of an online nature is necessary or whether we can rely on offline batch computation instead. Our analyses of a toy data set show that Bayesian online method is preferred.

3.1 Cumulative sum analysis

CUSUM records the running difference between the samples from a process x_t for $t = 1, ..., n$ and the average $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$:

$$\text{CUSUM} = \begin{cases} S_0 = 0 \\ S_t = S_{t-1} + (x_t - \bar{x}) \end{cases} \tag{3.1}$$

Segments of the plotted chart with a downward slope signal improvement in skills, as this indicates where the values are below the average. This algorithm is of an offline nature and is run only once all the time series data had been measured and its recursive workflow is shown in Algorithm 1. Mathematically speaking the inflection point of a learning curve is described as being a cubic polynomial attaining its global maximum point. It may also be the case that the inflection point is observed at a global minima. In that instance, the learning curve is concave upward past the inflection point, i.e. time series is at a phase which is above

the process mean and the CUSUM is increasing. These instances can occur when there is no learning curve observed or when the global maxima is observed elsewhere outside the range of studied data points.

Algorithm 1: Cumulative sum analysis

1: **Initalise score with mean value:**

$$S_0 = 0$$
$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

2: **Observe a new datum** x_t**.**

3: **Compute and update score:**

$$S_t = S_{t-1} + (x_t - \bar{x})$$

4: **Return to Step 2.**

3.1.1 Synthetic data example

Using $T = 1000$ randomly generated data points we show the output of using the CUSUM method in Figure 3.1 and show the limitations of using this method for the learning curves used by [30] in Figures 3.2, 3.3. Let us therefore suppose that there exist some task which requires many hours to hone the skills of, for instance, learning a foreign language [44]. There is an initial period of steady improvement in someones language skills, followed by a period of limited progress caused due to confusion of learning new conjugation of verbs, before the person is finally able to write a short essay in that language.

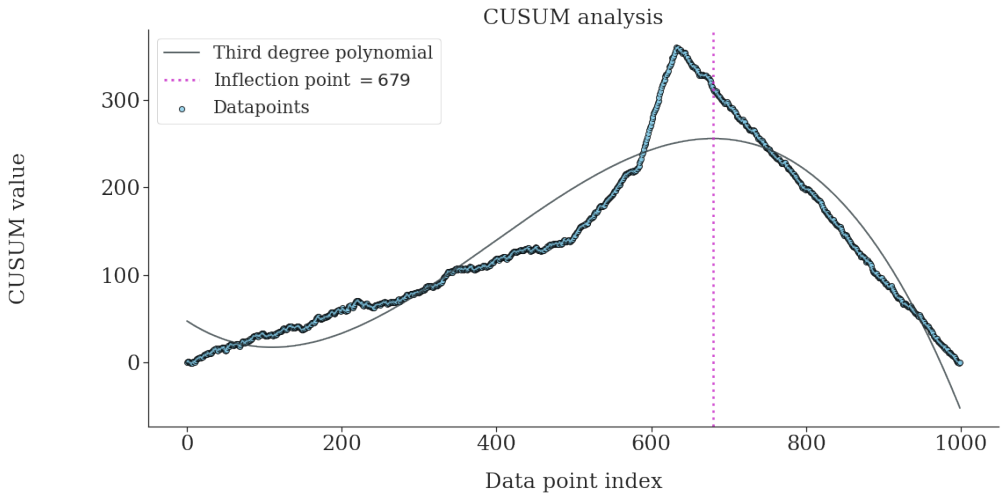


Figure 3.1: CUSUM analysis for $T = 1000$ randomly generated data points. The inflection point is marked with a dashed red line and is depicted on its global maximum point $t = 679$.

Within surgery setting similar rules are applied. For a newly graduated medical student who had never before performed surgery in a real life setting, progression can be expected to be slow at the start and hence the learning curve would be on the rise. This is due to using the new Mako machinery requires a lot of guidance from the head surgeon. After a little while the surgeon will be expected to perform surgery independently but this may result in longer surgery times due to the meticulousness factor of wanting to

avoid mistakes. Finally however, the surgeon would had performed enough surgeries to gain confidence and see their learning curve slope downward. We plot this hypothetical scenario with all three discussed phases in Figure 3.1.

3.1.2 Caveats for interpretation

Three main issues arise when it comes to using the CUSUM method for the analysis of a surgeons skills in using the Mako RAS system. Firstly, the number of cases a surgeons attends to will affect where and if a learning curve is to be found. Secondly, any previous training, be it in conventional surgery or with another robot, will affect the learning curve and thus prevents straightforward comparison between surgeons. Thirdly, the basic assumption here is that only two learning phases exist for surgical competence. There are copious real world examples where this is incorrect. Think back to the language example given earlier or to a judoka athlete obtaining different belts. These are particularly problematic from a business stand point for a clinic that wants to incorporate operating theater time inside a business plan. A clinic would be much better placed if it did not expect each surgeon to attend to the same number of cases but instead had a robust way of monitoring their individual progress.

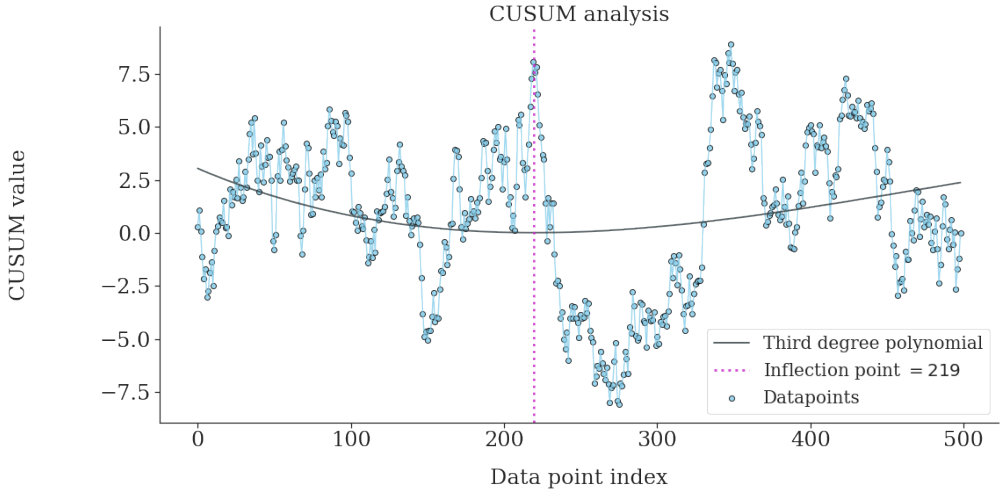


Figure 3.2: CUSUM analysis for the first period $T_{t=1:500} = 500$ of randomly generated data points. The inflection point is marked with a dashed red line and is depicted on its global minima point $t = 219$.

Observe how the case level is a major factor in Figures 3.1 - 3.3 where the data is all drawn from the same pool of hypothetical cases but display drastically different results if taken independently of each other. For an insufficient number of cases a learning curve is not found in Figure 3.2 as the inflection point is a global minima because the trend is still rising. This can be attributed to the surgeon still learning and having an insufficient number of cases to show for. However, if we take the same number of cases but for a senior surgeon who had already completed their initial training phase then the learning curve in Figure 3.3 looks completely different with an inflection point found to be in a global maxima.

Clearly then stating that no learning curve exists for some surgical task based only on what we had observed in Figure 3.2 would be inaccurate. Similarly, stating that surgeon A is faster than surgeon B in learning how to operate the MAKO RAS robot does not paint the complete picture. Instead, we should take into account the surgeons level of expertise and only then assess the skills of surgeons accordingly. Furthermore, assuming that a surgeon had reached their pinnacle of competence in operating the MAKO robot through traditional observation of a learning curve is a hasty conclusion because it detracts from the idea that medicine is a constantly evolving field. Until a golden standard in surgery is found, it is more reasonable to assume that there exist extra room for improvement with additional cases.

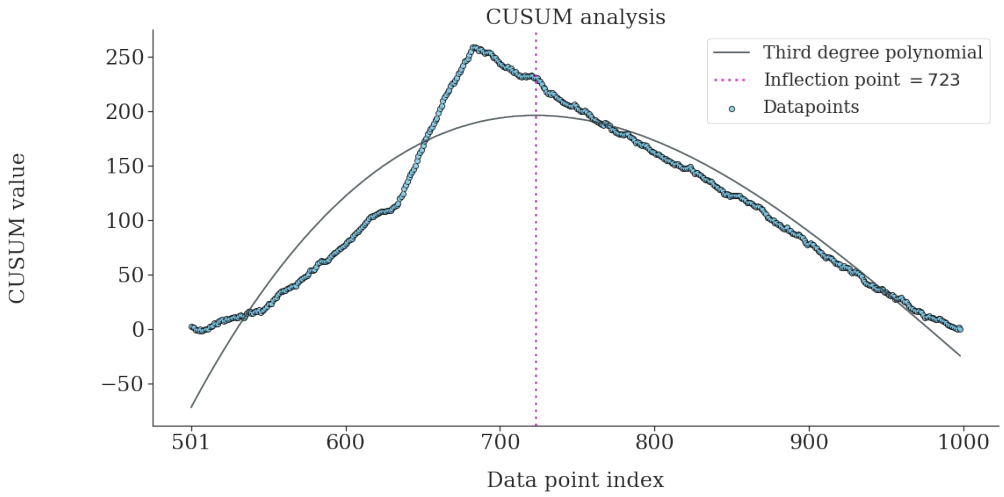


Figure 3.3: CUSUM analysis for the second period $T_{t=501:1000} = 500$ of randomly generated data points. The inflection point is marked with a dashed red line and is depicted on its global maximum point $t = 723$.

3.2 Offline change point detection

The second offline algorithm that we examine is the one by Fearnhead [33, 34] which we will refer to as the offline BCD. This algorithm again requires for all the time series data to first be recorded and aims to sample from the posterior distribution of change point locations. Recursions inside this algorithm borrow heavily from the Forward-Backward algorithm, otherwise known as Baum-Welch [45], which updates the parameters inside of a hidden Markov model (HMM) using expectation maximisation (EM). The Forward-Backward algorithm computes the probability of being in a particular state at a given time, given a sequence of observations. This probability is computed using a combination of forward and backward messages, which leverages dynamic programming to be recursively computed.

In the context of the Forward-Backward algorithm, the EM approach consists of two steps: the *E*-step and the *M*-step [39, p.607-625]. In the *E*-step, the sufficient statistics of the hidden variables are computed, given the current estimate of the parameters. In the *M*-step, the parameters are updated to maximise the expected likelihood of the observed data, based on the expected sufficient statistics computed in the *E*-step. The EM algorithm can then be used to estimate the transition probabilities and emission probabilities of the HMM, based on the observed sequences and the probabilities computed using the Forward-Backward algorithm. In the context of surgical procedure and for each surgical task we take the observed time. We then compute the transitions between phases by modelling the probabilities of moving from one phase to the next. Monitoring the probabilities of each unobserved hidden state over time, it becomes possible to predict when a change point in the procedure is likely to occur, indicating a transition to a new phase of the surgery.

Fearnhead introduced slight variation for the offline BCD by computing from the end of a sequence of observations and making use of dynamic programming to update the parameters in the opposite direction. To best quote Fearnhead [34]:

> The assumption of independence between segments ensures the necessary Markov property that is required for Forward-Backward type recursions. For a data set consisting of observations at discrete times $1, ..., n$ the recursions are based on calculating the probability of the data from time t to time n, given a changepoint at time t, in terms of the equivalent probabilities at times $t + 1, ..., n$. Once these probabilities have been calculated for all time instances, it is possible to directly simulate from the posterior distribution of the time of the first changepoint, and then the conditional distribution of the time of the second changepoint, given the first, and so on.

It achieves this by taking the parameter values to find the posterior distribution of the latent variables, which in turn are used to evaluate the expectation of the likelihood function on the data in the first step. Since our aim is to sample from the posterior distribution and to make use of the EM algorithm in estimating the parameters of a model with hidden surgical phases, the algorithm simply leverages Bayesian inference tools provided in Section 2.2 and uses sufficient statistics to compute the MLE for those said parameters [39, p.615-618] in the second step.

We borrow the notations used in Fearnhead [33] with which to introduce the offline BCD algorithm. For n independent variables $\mathbf{x} = \{x_1, ..., x_n\}$ and m changepoints $\boldsymbol{\tau} = \{\tau_1, ..., \tau_m\}$, with $\tau_0 = 0$ and $\tau_{m+1} = n$, the data belonging to the k^{th} segment is assumed to be independent with the data on other segments given a set of parameters θ_k associated with the k^{th} segment for $k = 1, ..., m+1$. Then the data point x_i drawn from a density belonging to the k^{th} segment is denoted here $f(x_i|\theta_k)$.

For this project we define the change point priors on the model to be specified by a probability mass function $g(t)$ for the time between two successive points. This is chosen because we assume we do not have prior knowledge for the exact number of changepoints. See Fearnhead [33] for a prior based on the number of changepoints. The cumulative distribution function between two points is then defined as:

$$G(t) = \sum_{i=1}^{t} g(i) \tag{3.2}$$

Which implies a prior distribution on the changepoints. The likelihood function of evaluating how well the data for times $s \geq t$ can fit in one segment is given by:

$$\mathbb{P}(t, s) = \mathbb{P}(x_{t:s}|t, s \text{ belong to the same segment}) \tag{3.3}$$

With $x_{t:s}$ denoting the sequence of all data points from t to s. We can now begin with the backward recursion part with defining the likelihood function on the observed data for $t = 2, ..., n$:

$$\begin{aligned}
Q(t) &= \mathbb{P}(x_{t:n}|\text{change point occurred at } t-1) \\
&= \sum_{s=t}^{n-1} \mathbb{P}(x_{t:n}, \text{next change point is at } s) + \mathbb{P}(x_{t:n}, \text{no further changepoints}) \\
&= \sum_{s=t}^{n-1} \mathbb{P}(\text{next change point is at } s) \cdot \mathbb{P}(x_{t:s}, x_{s+1:n}|\text{next change point is at } s) \\
&\quad + \mathbb{P}(x_{t:n}|t, n \text{ belong to the same segment}) \cdot \mathbb{P}(\text{segment length is} > n-t) \\
&= \sum_{s=t}^{n-1} g(s+1-t) \cdot \mathbb{P}(x_{t:s}|t, s \text{ belong to the same segment}) \cdot \mathbb{P}(x_{s+1:n}|\text{change point is at } s) \\
&\quad + \mathbb{P}(x_{t:n}|t, n \text{ belong to the same segment}) \cdot (1 - G(n-t)) \\
&= \sum_{s=t}^{n-1} g(s+1-t) \cdot \mathbb{P}(t, s) \cdot Q(s+1) + \mathbb{P}(t, n) \cdot (1 - G(n-t))
\end{aligned} \tag{3.4}$$

Where we drop the notation conditioning on a change point occurring at $t-1$ for convenience. Similarly, $Q(1) = \mathbb{P}(x_{1:n})$ since the series begins with a change point on $\tau_0 = 0$ by default. Once we iterate over all the data backwards for $t = n, .., 1$, the forward recursion for the inference of changepoints take place where the posterior distribution of the first change point τ_1 is given by:

$$\mathbb{P}(\tau_1|x_{1:n}) = \frac{\mathbb{P}(x_{1:n}, \tau_1)}{\mathbb{P}(x_{1:n})} \tag{3.5}$$

$$= \frac{\mathbb{P}(\tau_1) \cdot \mathbb{P}(x_{1:\tau_1}|\tau_1) \cdot \mathbb{P}(x_{\tau_1+1:n}|\tau_1)}{Q(1)} \tag{3.6}$$

$$= \frac{\mathbb{P}(1, \tau_1) \cdot Q(\tau_1 + 1) \cdot g(\tau_1)}{Q(1)}$$

For $\tau_j = \tau_{j-1} + 1, ..., n - 1$. Since there are m changepoints and $\tau_{m+1} = n$ is a change point when the time series terminates, the posterior probably of observing no changepoints is:

$$\mathbb{P}(\tau_{m+1}|x_{1:n}) = \frac{\mathbb{P}(1, n)(1 - G(n - 1))}{Q(1)} \tag{3.7}$$

In fact, for any future change point τ_j having observed a change point at a previous time step τ_{j-1} the posterior probability is simply:

$$\mathbb{P}(\tau_j|\tau_{j-1}, x_{1:n}) = \frac{\mathbb{P}(\tau_{j-1} + 1, \tau_j)Q(\tau_j + 1)g(\tau_j - \tau_{j-1})}{Q(\tau_{j-1} + 1)} \tag{3.8}$$

Whilst the probability of observing no further changepoints is:

$$\mathbb{P}(\tau_{m+1}|\tau_{j-1}, x_{1:n}) = \frac{\mathbb{P}(\tau_{j-1} + 1, n)(1 - G(n - \tau_{j-1} - 1))}{Q(\tau_{j-1} + 1)} \tag{3.9}$$

A sequential and complete implementation of the Forward-Backward method is presented in Algorithm 2. In summary, we use Bayes' Rule from Theorem 1 to compute the posterior distribution over the possible change point locations given the data. This involves multiplying the prior distribution by the likelihood function and normalising to obtain a probability distribution. The expectation maximisation for the locations of the changepoints is then repeated to maximise the posterior probability.

In the univariate case, by Equation 61 in Murphy [46] we know that for data from a univariate Gaussian with unknown parameters $\theta = \{\mu, \sigma^2\}$ the likelihood function is Normal:

$$\mathbb{P}(\mathbf{x}|\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2) \tag{3.10}$$

With the conjugate prior on parameters θ being from a Normal-Inverse-Gamma (NIG) distribution:

$$\begin{aligned} \mathbb{P}(\mu, \sigma^2) &= \mathrm{N}\Gamma^{-1}(\mu, \sigma^2|\mu_0, \kappa_0, \alpha_0, \beta_0) \\ &= \mathcal{N}(\mu|\mu_0, \kappa\sigma^2) \cdot \Gamma^{-1}(\sigma^2|\alpha_0, \beta_0) \end{aligned} \tag{3.11}$$

Then due to conjugacy the posterior is also from an NIG:

$$\begin{aligned} \mathbb{P}(\mu, \sigma^2|\mathbf{x}) &= \mathbb{P}(\mu, \sigma^2) \cdot \mathbb{P}(\mathbf{x}|\mu, \sigma^2) \\ &= \mathrm{N}\Gamma^{-1}(\mu, \sigma^2|\mu_0, \kappa_0, \alpha_0, \beta_0) \cdot \mathcal{N}(\mu, \sigma^2) \end{aligned} \tag{3.12}$$

The posterior predictive of a new data point is then:

$$\mathbb{P}(x^{\text{new}}|\mathbf{x}, \mu, \sigma^2) = \mathrm{N}\Gamma^{-1}(\mu, \sigma^2|\mu_n, \kappa_n, \alpha_n, \beta_n) \tag{3.13}$$

Figure 3.4 demonstrates the offline BCD applied on the $T = 1000$ randomly generated data points used in Section 3.1.1. The posterior probability over changepoints is modelled with a Student's t distributed likelihood model $\mathbb{P}(x_{t:s}|\mu_k, \kappa_k, \alpha_k, \beta_k)$ where $\mu_0 = 0$, $\alpha_0 = \beta_0 = \kappa_0 = 1$. The hyperparameters are updated as in Sholihat et al. [37]:

$$\mu_n = \frac{\kappa_{n-1} \cdot \mu_{n-1} + x_{t:s}}{\kappa_{n-1} + 1} \tag{3.14}$$

$$\sigma_n = \sqrt{\frac{2\beta_n \cdot (\kappa_{n-1} + 1)}{\kappa_{n-1}}} \tag{3.15}$$

Algorithm 2: Offline Bayesian change point detection

1: **Define the prior distribution over the parameters:**

$$\theta_0 = \mu, \sigma \sim \mathrm{N}\Gamma^{-1}(\mu, \sigma|\mu_0, \alpha_0, \kappa_0, \beta_0)$$

2: **Compute cumulative distribution function:**

$$G(n) = n \cdot \lambda$$

3: **Initalise data sequence and likelihood of data:**

$$\begin{aligned} \mathbb{P}(n-1, n) &= f(x_n|\theta_0) \\ Q(n) &= f(x_n|\theta_0) \end{aligned}$$

4: **Evaluate recursively backwards the data between two points $i < j \leq n-1$:**

$$\mathbb{P}(i, j) = f(x_{i:j}|\theta_k)$$

5: **Compute the likelihood function on the observed data:**

$$Q(i) = \mathbb{P}(i, j) \cdot Q(j) \cdot g(j-i) + \mathbb{P}(i, j) \cdot (1 - G(j-i))$$

6: **Compute probability of next change point:**

$$\mathbb{P}(\tau_i|\tau_j, x_{1:n}) = \frac{\mathbb{P}(i, j)Q(j)g(j-i)}{Q(i)}$$

7: **Update distribution function parameters for the n^{th} step.**

$$\begin{aligned} \kappa_{n+1} &= \kappa_n + 1 \\ \mu_{n+1} &= \frac{\kappa_n \cdot \mu_n + x_{1:n}}{\kappa_n} \\ \alpha_{n+1} &= \alpha_n + 0.5 \\ \beta_{n+1} &= \beta_n + \frac{\kappa_n \cdot (x_{1:n} - \mu_n)^2}{2(\kappa_n + 1)} \end{aligned}$$

8: **Return to Step 4 while $0 < i < j$.**

9: **Compute posterior distribution over the change point using forward recursion:**

$$\mathbb{P}(\tau_j|\tau_i, x_{1:n}) = \frac{\mathbb{P}(i, j)Q(j)g(j-i)}{Q(i)}$$

10: **Return to Step 9 while $i < j < n$.**

$$\nu_n = 2\alpha_n \tag{3.16}$$

Where the hyperparameter α_k is used to define the new degrees of freedom and β_k scales the variance factor:

$$\kappa_n = \kappa_{n-1} + 1 \tag{3.17}$$

$$\alpha_n = \alpha_{n-1} + 0.5 \tag{3.18}$$

$$\beta_n = \beta_{n-1} + \frac{\kappa_{n-1} \cdot (x_{t:s} - \mu_{n-1})^2}{2(\kappa_{n-1} + 1)} \tag{3.19}$$

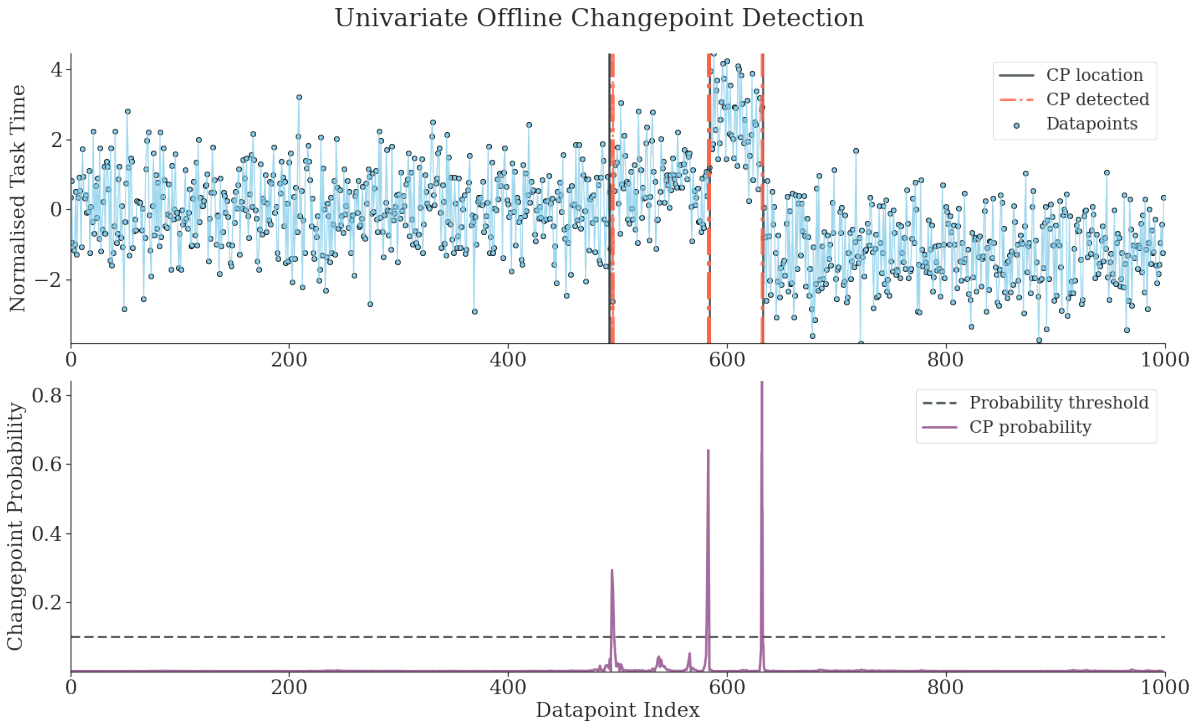


Figure 3.4: Offline change point detection for $T = 1000$ synthetic data points modelled from a Gaussian distribution. The marginal likelihood $\mathbb{P}(x_{t:s}|\theta) \sim t(x_{t:s}|\mu_n, \kappa_n, \alpha_n, \beta_n)$ that data from t to s is produced by a single model θ_k is modelled from a Student's t distributed likelihood model.

The top plot in Figure 3.4 shows the normalised time series data together with three preset and detected change point locations. The bottom plot depicts the posterior change point probability $\mathbb{P}(\tau_j|\tau_{j-1}, x_{1:n})$ with a probability threshold of 0.1. Algorithm 2 distinguishes between the four different segments and assigns high probability to each detected change point location.

3.2.1 Multivariate likelihood models

Often in surgery there are several data rich surgical tasks to assess. The risk of modelling every data stream as an individual univariate model is that it ignores the covariance between non-stationary time series and thus is unable to capture correlations between the features. Xuan et al. [47] provides a full covariance model approach for modelling the likelihood of data sequence belonging to a single model segment. The proof follows the extensive Bayesian analysis work found in Section 3.4 and Section 9.5 of Murphy [46].

For a multi-parameter model we therefore employ the multivariate Normal distribution:

$$\mathbb{P}(\mathbf{x}_p|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.20}$$

The conjugate prior is from a Normal-Inverse-Wishart distribution:

$$\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{NIW}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{\mu}_0, \kappa_0, \nu_0, \boldsymbol{\Sigma}_0^{-1}) \tag{3.21}$$

With ν degrees of freedom. It can then be shown that the posterior predictive distribution over a new data point belongs to a t distribution for p dimensions:

$$\mathbb{P}(x^{\text{new}} | \mathbf{x}_p, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{\nu_n + p}{2})}{\Gamma(\frac{\nu_n}{2}) \nu_n^{\frac{p}{2}} \pi^{\frac{p}{2}} |\boldsymbol{\Sigma}_n|^{\frac{1}{2}}} \left[1 + \frac{1}{\nu_n} (\mathbf{x}_p - \boldsymbol{\mu}_n)^\intercal \boldsymbol{\Sigma}_n^{-1} (\mathbf{x}_p - \boldsymbol{\mu}_n) \right]^{-\frac{\nu_n + p}{2}} \tag{3.22}$$

We demonstrate the difference in methods between the independent features and full covariance models in Figure 3.5 where the time series data is modelled from Gaussian distributions with $p = 3$. Both methods are viable candidates for offline change point detection in multivariate time series and are able to identify both change point locations at 90 and 178. The full covariance model is able to outperform the independent features model by assigning a larger change point probability at each location because it captures the correlations between all three data sequences.

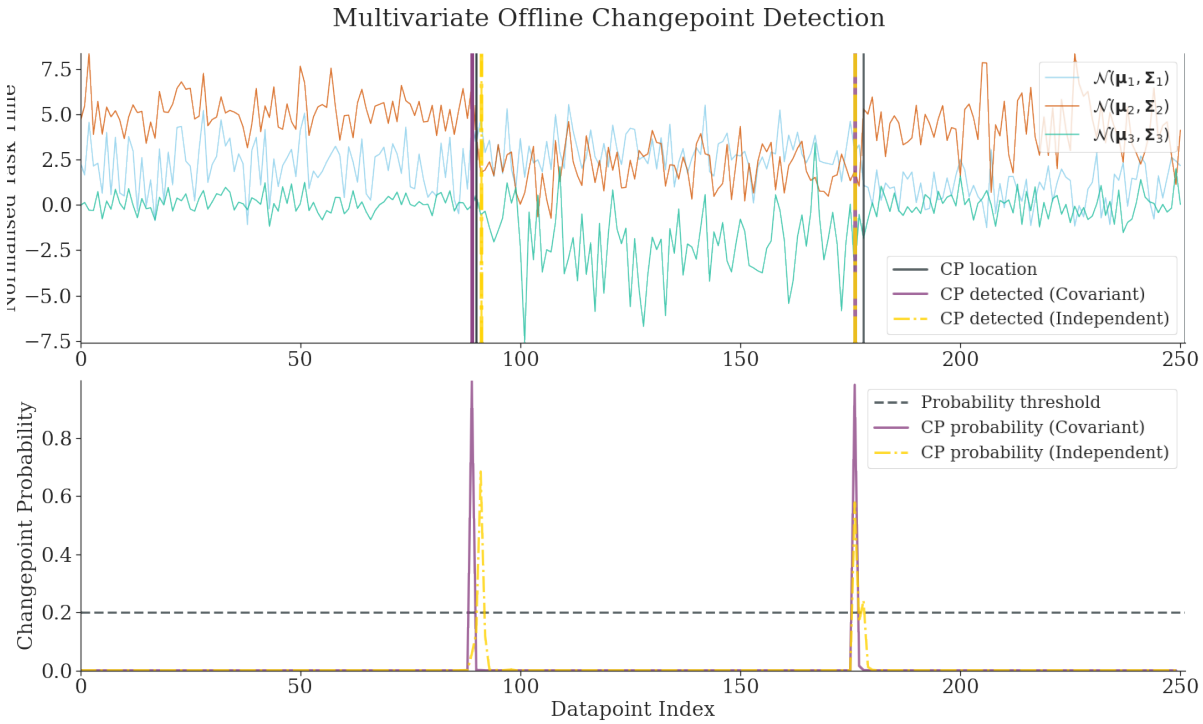


Figure 3.5: Multivariate offline change point detection with and without covariance factor for $T = 250$ synthetic data points modelled from three Gaussian distributions.

The full covariance model clearly outperforms the independent factor model in the event of 15 change point locations in Figure 3.6. The former model assigns higher change point probability between each partition, with the two most interesting changepoints taking place at 1151 and 2193. On both instances the Gaussian distributions $\mathcal{N}(\mu, \boldsymbol{\Sigma})$ have very different segment mean and standard deviation from the mean, yet the independent factor model is unable to pick a change point at those locations.

3.3 Bayesian online change point detection

Our aim is to infer the posterior predictive distribution $\mathbb{P}(x_{t+1} | \mathbf{x}_{1:t})$. This algorithm is of an online nature as we are conditioning on past observations with a constant inflow of new data. Inference about particular

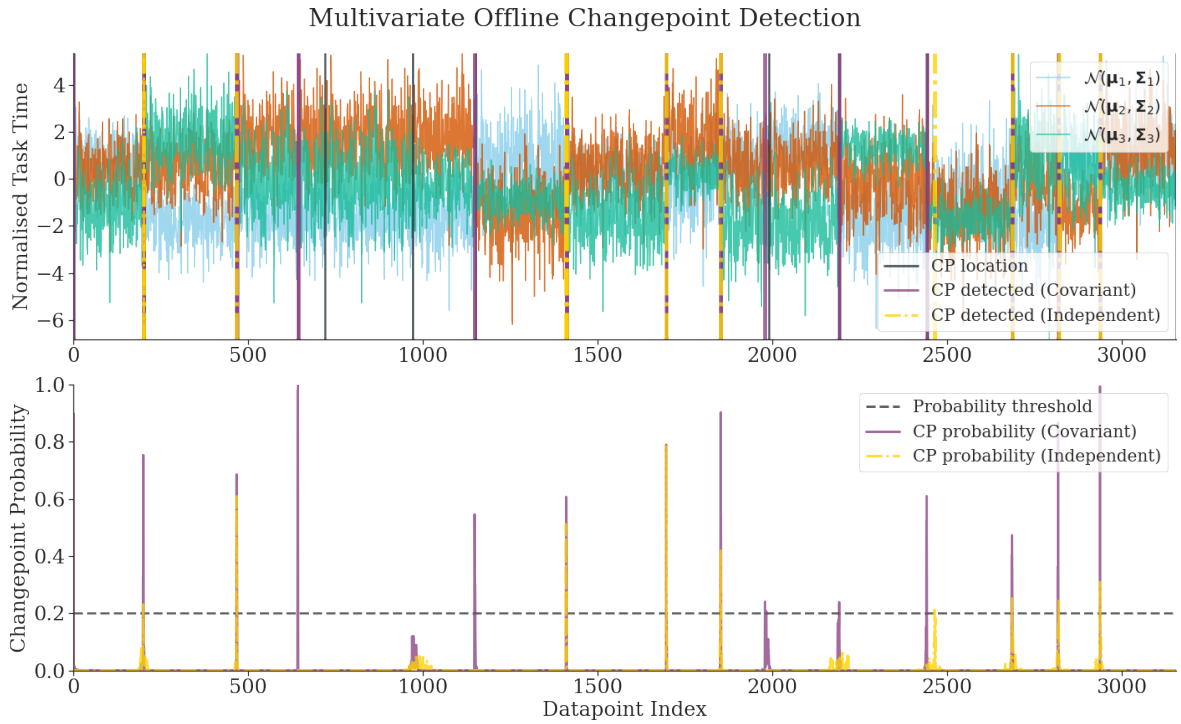


Figure 3.6: Multivariate offline change point detection with and without covariance factor for $T = 3156$ synthetic data points modelled from three Gaussian distributions with 15 change point partitions.

future observation may then be made by first conditioning on the partition over the current run length r_t at time t. What we then do is assume that observations in different partitions of the data are independent and deduce the posterior distribution $\mathbb{P}(r_t|\mathbf{x}_{1:t})$ over the current run length inside of the sequence $\mathbf{x}_{1:t}$. This is achieved by generating a distribution of the next unseen data point x_{t+1}, given that we had observed data points $x_s, x_{s+1}, ..., x_{t-1}, x_t$ for $s \leq t$, using a message passing system.

The application of both these posterior distributions inside BOCD is made possible using Bayesian methods for conditional probabilities and marginalisation that we had encountered in Section 2.1. The goal is to recursively update the run length estimation with every new data point. Then for BOCD, the posterior predictive probability of the next data point x_{t+1}, given all the observations so far $\mathbf{x}_{1:t}$, is computed by taking Equation 2.6 to sum over every possible parameter value of x_{t+1} that has a joint distribution with $\mathbf{x}_{1:t}$ and then marginalising over the run length r_t:

$$\begin{aligned}
\mathbb{P}(x_{t+1}|\mathbf{x}_{1:t}) &= \frac{\mathbb{P}(x_{t+1}, \mathbf{x}_{1:t})}{\mathbb{P}(\mathbf{x}_{1:t})} \\
&= \frac{\sum_{r_t} \mathbb{P}(x_{t+1}, r_t, \mathbf{x}_{1:t})}{\mathbb{P}(\mathbf{x}_{1:t})} \\
&= \frac{\sum_{r_t} \mathbb{P}(x_{t+1}|r_t, \mathbf{x}_{1:t})\mathbb{P}(r_t|\mathbf{x}_{1:t})\mathbb{P}(\mathbf{x}_{1:t})}{\mathbb{P}(\mathbf{x}_{1:t})} \\
&= \sum_{r_t} \mathbb{P}(x_{t+1}|r_t, \mathbf{x}_{1:t})\mathbb{P}(r_t|\mathbf{x}_{1:t}) \\
&= \sum_{r_t} \mathbb{P}(x_{t+1}|r_t, \mathbf{x}_t^r)\mathbb{P}(r_t|\mathbf{x}_{1:t})
\end{aligned} \tag{3.23}$$

We denote $\mathbf{x}_t^r$ to be the data points covered only by the current run length r_t which contribute to the prediction of x_{t+1}. The interpretation here is that as new evidence flows in it does not determine our beliefs, but rather updates our *prior* beliefs. Thus we arrive at Equation 1 from the paper [35]. It takes in

essence the form of the predictive distribution showcased in Equation 8.24 from [41, p.268] over a set of independently and identically distributed training data $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ and model parameters θ that govern the distribution of $\mathbf{x}$:

$$\mathbb{P}(x^{\text{new}}|\mathbf{x}) = \int \mathbb{P}(x^{\text{new}}|\theta)\mathbb{P}(\theta|\mathbf{x})\ \mathrm{d}\theta \tag{3.24}$$

To recover the density of the next data point x^{new}, the probabilistic model $\mathbb{P}(x^{\text{new}}|\theta)$ weighs itself by the posterior predictive distribution $\mathbb{P}(\theta|\mathbf{x})$ since it accounts for uncertainty about θ, something that MLE does not do. To infer a solution from Equation 3.23, it is assumed we can iteratively compute the predictive distribution that is conditional on some run length r_t. All that is left to find is the run length posterior distribution:

$$\mathbb{P}(r_t|\mathbf{x}_{1:t}) = \frac{\mathbb{P}(r_t, \mathbf{x}_{1:t})}{\mathbb{P}(\mathbf{x}_{1:t})} \tag{3.25}$$

We again make use of marginalisation to compute the joint distribution inside the numerator:

$$\begin{aligned}
\mathbb{P}(r_t, \mathbf{x}_{1:t}) &= \sum_{r_{t-1}} \mathbb{P}(r_t, r_{t-1}, \mathbf{x}_{1:t}) \\
&= \sum_{r_{t-1}} \mathbb{P}(r_t, r_{t-1}, x_t, \mathbf{x}_{1:t-1}) \\
&= \sum_{r_{t-1}} \mathbb{P}(r_t, x_t|r_{t-1}, \mathbf{x}_{1:t-1})\mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1}) \\
&= \sum_{r_{t-1}} \mathbb{P}(x_t|r_{t-1}, \mathbf{x}_{1:t-1})\mathbb{P}(r_t|r_{t-1}, \mathbf{x}_{1:t-1})\mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1}) \\
&= \sum_{r_{t-1}} \mathbb{P}(x_t|r_{t-1}, \mathbf{x}^r_{t-1})\mathbb{P}(r_t|r_{t-1})\mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1})
\end{aligned} \tag{3.26}$$

The term $\mathbf{x}_{1:t-1}$ drops out from inside the change point prior $\mathbb{P}(r_t|r_{t-1})$ because we base our assumption on the probability of a change point occurring at a given time t. This can be expressed using some intensity value taken from a distribution of previously observed run length. Similarly, the term r_t drops out of the predictive distribution over the newly observed data point x_t because this depends only on the data since the last change point. Lastly, the joint distribution $\mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1})$ on the previous run lengths probabilities gives this algorithm its iterative charm and acts as the message passing term.

All that is left for us to do is initialise a change point probability and to set some priors to feed into the underlying probabilistic model $\mathbb{P}(x_{t+1}|r_t, \mathbf{x}^r_t)$, just as we had done in Equation 2.44. Then similar to the authors in [35] we assume that a change point has occurred at the initial starting point, forcing the probability mass for the initial run length to zero $\mathbb{P}(r_0 = 0) = 1$, and that the run length either continues to grow with $r_t = r_{t-1} + 1$ or a change point has occurred at $r_t = 0$:

$$\mathbb{P}(r_t|r_{t-1}) = \begin{cases} H(r_{t-1} + 1) & \text{if } r_t = 0 \\ 1 - H(r_{t-1} + 1) & \text{if } r_t = r_{t-1} + 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.27}$$

Not only does this add simplicity to the model since we only have to worry about two cases with probability mass to compute, but we can also compute the *hazard function* once by setting it to be a constant $H(\cdot) = p$ with some probability of success p. Since the initial run length at the start is zero with probability one, the associated hyperparameters of this particular run length are simply the priors:

$$\nu^0_1 = \nu_{\text{prior}} \tag{3.28}$$

$$\boldsymbol{\chi}_1^0 = \boldsymbol{\chi}_{\text{prior}} \tag{3.29}$$

Assuming that the model is from an exponential family exactly as it was shown in Section 2.2.5, we are set and ready to iteratively solve the BOCD algorithm. Algorithm 3 compiles these computations together as we are now equipped to find the marginal predictive distribution in Equation 3.23.

To better understand the BOCD algorithm we provide an illustration in Figure 3.7 which is adapted from [35]. In the left hand side figure the seven data points $x_{i=\{1:7\}}$ belong to two partitions $p_{i=\{1,2\}}$ separated by one changepoint on the mean that occurs in the interval $t = \{3, 4\}$. The right hand side figure shows the run length r_t as a function of time t passing its probability mass along the solid lines. When a changepoint occurs, $r_4 = 0$ drops to zero, otherwise run length increases by one. The dashed lines indicate the possibility of a run length being truncated after a changepoint. In this example we are aware of when each change point occurs. During other scenarios it is not possible to explicitly know when the run length drop down to zero. However, the message passing term will continue travelling along this *trellis* since we are dealing with probabilities from the joint distribution, that is why the dashed lines are also important to consider here. Note that these message passing terms do not only travel along the diagonal for illustrative purposes only, but these also have a strong connection to how the algorithm works numerically.

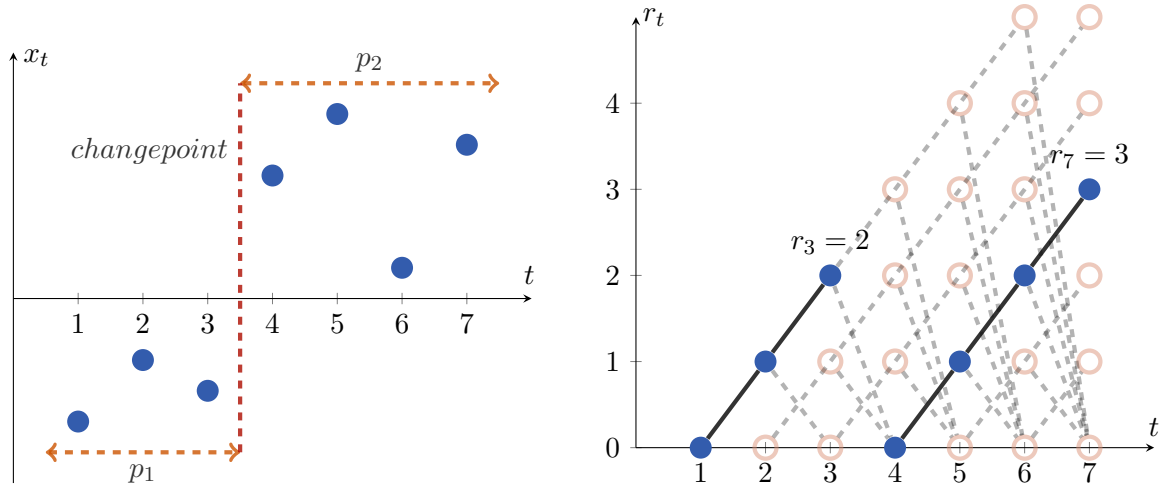


Figure 3.7: Changepoint model expressed in terms of run lengths. Figure adapted from [35].

We aim to further improve the understanding surrounding how Algorithm 3 deals with the message passing term numerically by showing how it is used in computing the posterior distribution over the current run length in Figure 3.8. The left hand side shows the run length posterior distribution with a logarithmic scale, where darker shades indicates higher probability of the run length increasing by one. These probabilities are computed along the diagonal just as we had illustrated for Figure 3.7. To obtain the posterior distribution at time t it is best to visualise the probabilities as a matrix of values, which is done in the right hand side figure. This is the exact same visualisation only mirrored, but in reality is the way most programming languages would populate this matrix.

Suppose then that we find ourselves in the middle column of the matrix in Figure 3.8 at time $t = 4$ (shown in dashed blue area) and we wish to follow the steps of Algorithm 3 in order to compute the posterior distribution. First we observe the new data point x_5. We then evaluate the predictive probability of x_5 under the posterior predictive distribution associated with the previous run length at time $t = 4$:

$$\pi_5^r = \mathbb{P}(x_5 | \nu_5^r, \boldsymbol{\chi}_5^r) \tag{3.30}$$

For $r = \{0, 1, 2, 3, 4\}$. The two posterior distributions are discussed in Section 3.4 and Section 3.5. We then compute two near identical steps. To obtain the growth probability for the run length increasing by a step size of one (shown in solid blue area) we simply make use of the bookkeeping being performed inside

Algorithm 3: Bayesian online change point detection

1: **Initalise change point and priors:**

$$\begin{aligned}\mathbb{P}(r_0 = 0) &= 1 \\ \nu_1^0 &= \nu_{\text{prior}} \\ \boldsymbol{\chi}_1^0 &= \boldsymbol{\chi}_{\text{prior}}\end{aligned}$$

2: **Observe a new datum x_t.**

3: **Compute the predictive probability of the underlying model over the current run length:**

$$\pi_t^r = \mathbb{P}(x_t | \nu_t^r, \boldsymbol{\chi}_t^r)$$

4: **Compute the growth probabilities for every possible run length value:**

$$\mathbb{P}(r_t = r_{t-1} + 1, \mathbf{x}_{1:t}) = \mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1}) \cdot \pi_t^r \cdot (1 - H(r_t - 1))$$

5: **Compute the change point probabilities for every run length dropping to zero:**

$$\mathbb{P}(r_t = 0, \mathbf{x}_{1:t}) = \sum_{r_{t-1}} \mathbb{P}(r_{t-1}, \mathbf{x}_{1:t-1}) \cdot \pi_t^r \cdot H(r_{t-1})$$

6: **Compute the evidence:**

$$\mathbb{P}(\mathbf{x}_{1:t}) = \sum_{r_t} \mathbb{P}(r_t, \mathbf{x}_{1:t})$$

7: **Compute the posterior distribution over the current run length:**

$$\mathbb{P}(r_t | \mathbf{x}_{1:t}) = \frac{\mathbb{P}(r_t, \mathbf{x}_{1:t})}{\mathbb{P}(\mathbf{x}_{1:t})}$$

8: **Update the sufficient statistics:**

$$\begin{aligned}\nu_{t+1}^0 &= \nu_{\text{prior}} \\ \boldsymbol{\chi}_{t+1}^0 &= \boldsymbol{\chi}_{\text{prior}} \\ \nu_{t+1}^{r+1} &= \nu_r^r + 1 \\ \boldsymbol{\chi}_{t+1}^{r+1} &= \boldsymbol{\chi}_t^r + \mathbf{u}(x_t)\end{aligned}$$

9: **Perform prediction of marginal predictive distribution from Equation 3.23**

$$\mathbb{P}(x_{t+1} | \mathbf{x}_{1:t}) = \sum_{r_t} \mathbb{P}(x_{t+1} | r_t, \mathbf{x}_t^r) \cdot \mathbb{P}(r_t | \mathbf{x}_{1:t})$$

10: **Return to Step 2.**

of the matrix by shifting the growth probabilites at the previous time step $t = 4$ down and to the right and multiplying with:

$$\mathbb{P}(r_5 = r_4 + 1, \mathbf{x}_{1:5}) = \mathbb{P}(r_4, \mathbf{x}_{1:4}) \cdot \pi_4^r \cdot (1 - H(r_4 - 1)) \tag{3.31}$$

The mere difference in computing the change point probability (shown in solid red area) is we sum over all possible values of the run length at time $t = 4$:

$$\mathbb{P}(r_4 = 0, \mathbf{x}_{1:4}) = \sum_{r_4} \mathbb{P}(r_4, \mathbf{x}_{1:4}) \cdot \pi_5^r \cdot H(r_4) \tag{3.32}$$

Of course we normalise across all values inside that particular column of the matrix before filling in the remainder rows of that column with zero values. These steps are repeated for the remainder of the columns before then plotting the results.

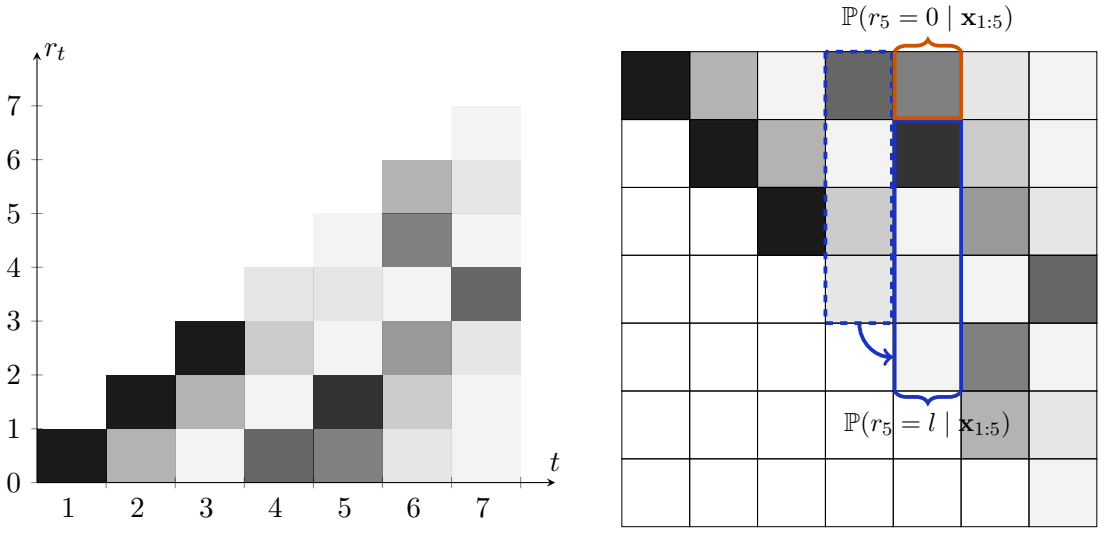


Figure 3.8: Run length posterior distribution $\mathbb{P}(r_t|\mathbf{x}_{1:t})$ for the changepoint model expressed as a matrix of probability values with a logarithmic scale. The left hand side graph is none other than a mirrored matrix that is used for bookkeeping.

3.4 Gaussian distributed posterior

Adams et al. [35] had presented three case studies which involved modeling of the data from a Gaussian distribution with unknown mean, Gaussian distribution with abrupt changes to the piecewise constant variance, and discrete data from a Poisson distribution. For our case on hand we forego the latter two cases. Firstly we can assume that a surgeon is not all of the sudden able to perform a surgery task in record time over night, but rather the learning process has a slow moving trend, hence there are no abrupt changes to the variance. Secondly since surgery task times are continuous we have no use of modelling the data using a Poisson distribution.

Instead, we model the posterior predictive probability $\mathbb{P}(x_{t+1}|\mathbf{x}_{1:t}) = \mathcal{N}(x|\mu_n, \sigma_n + \sigma)$ as a Gaussian process with unknown mean μ and a known variance σ^2, whilst also incorporating priors μ_0, σ_0^2. This assumption that the mean is unknown but the variance is within some range follows from Section 2.2.2. The parameters μ_n, σ_n^2 are none other than $\mu_{\text{MAP}}, \sigma_{\text{MAP}}$ after n data points had been observed. See Section 2.2.2 for detail. The unknown mean $\mu \sim \mathcal{N}(\mu|\mu_0, \sigma_0^2)$ changes according to the priors.

The run length probability together the probability of change points are shown in Figure 3.9. We make use of normalising the run length probabilities $\mathbb{P}(r_t, \mathbf{x}_{1:t})$ for improved numerical stability. The original

method in [35] is a bit difficult to make use of to the untrained eye in an industry setting due to the pixelated regions of probability. Therefore we introduce a clear cut method for surgeons and clinical staff to understand whether a change point had taken place.

Firstly, we set the segment length l_t to be our assumption of how many data points are required before change point is detected. Note that the run length probability is a matrix of values, with each run length traveling diagonally upwards. Setting $l_t = x$ is equivalent to observing the run length probability for row x (from the bottom) of this matrix. Secondly, we calculate the average run length probability of row x. If two neighbouring pixels are on either side of the average probability, we mark the column indices of these probability values as the inversion points. Thirdly, by setting a probability threshold p_t and seeing which column values exceed this then signals the occurrence of a change point.

To showcase this method at work we generated $T = 1000$ random data points across four different segments with varying mean. This is the exact same synthetic data used in Section 3.1.1 to study learning curves using the CUSUM analysis method. The three change points for this dummy test data are set by us so to assess whether the algorithm is able to detect those. The hazard function of a change point occurring at time t was set to $H(\cdot) = 4 \times 10^{-3}$, with the prior parameters being $\mu_0 = 0$, $\sigma_0^2 = 2$ and the known variance of the data $\sigma^2 = 1$.

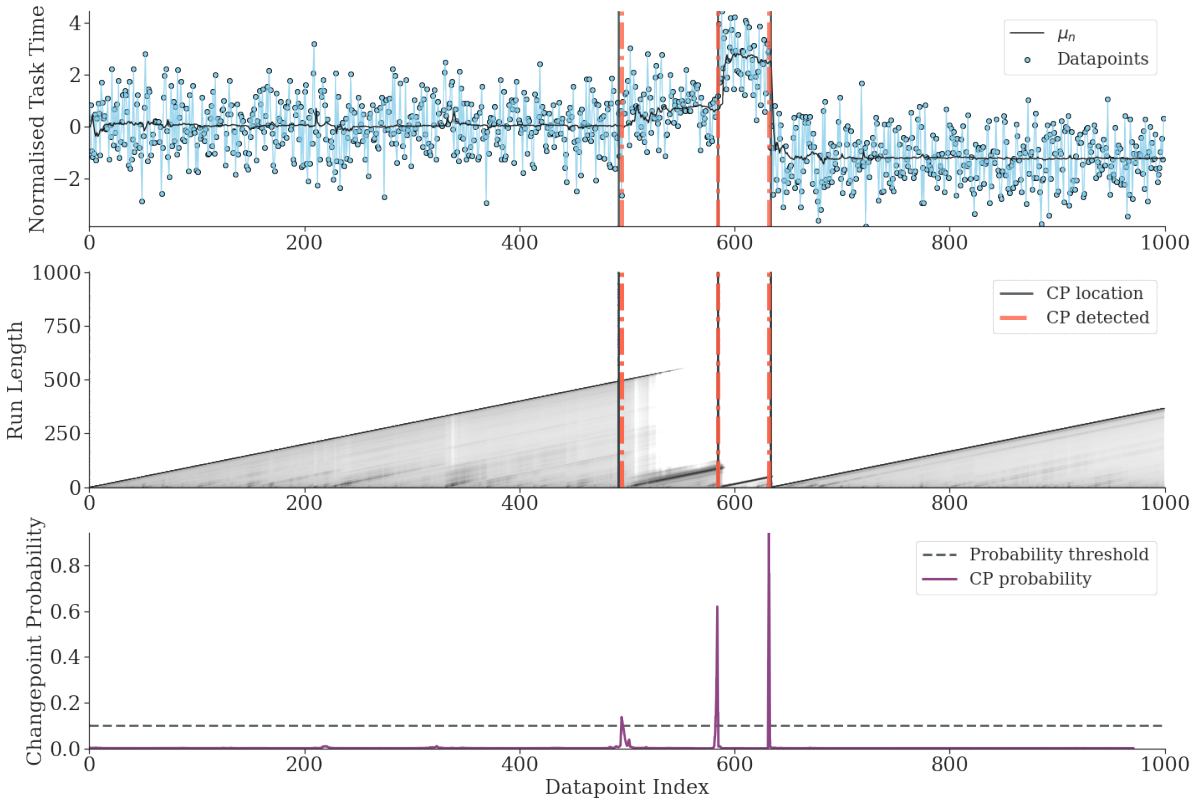


Figure 3.9: Bayesian online change point detection for $T = 1000$ synthetic data points modelled from a Gaussian distribution. The growth probability of run length $\mathbb{P}(r_t, \mathbf{x}_{1:t})$ modelled from a Gaussian distribution is able to pick up the three change points with almost near perfection.

The top plot in Figure 3.9 shows the normalised values over time, with the data points normalised around 0. At each time step t, the predictive mean $\hat{\mu}_t$ is modelled from a Gaussian distribution and is plotted using a solid black line. The middle plot shows the posterior probability of the current run length $\mathbb{P}(r_t|\mathbf{x}_{1:t})$ at each time step using a logarithmic colour scale. Darker pixels indicate higher probability of a run r_t of length t emerging. The vertical black lines indicate that we set a change point to take place inside the dummy test data, whilst the red dotted lines suggest that the BOCD algorithm detected a change point. The bottom

plot tracks the run length probability of each data point belonging to a segment of length $l_t = 50$, whilst the probability threshold was set to $p_t = 0.1$.

3.5 Student's t distributed posterior

In the event that a new surgeon had started and completed only a handful of surgeries, the sample size may be too small to model the BOCD using a Gaussian process. We therefore also investigate the novel case where both the mean and variance are unknown and are distributed according to a Normal-gamma distribution $\mu, \sigma \sim \text{N}\Gamma(\mu, \sigma|\mu_0, \alpha_0, \kappa_0, \beta_0)$. The posterior predictive probability is then t-distributed $\mathbb{P}(x_{t+1}|\mathbf{x}_{1:t}) = t_{2\alpha_n}\left(x|\mu_n, \frac{\beta_n(\kappa_n+1)}{\alpha_n\kappa_n}\right)$ with mean μ_n, variance $\frac{\beta_n(\kappa_n+1)}{\alpha_n\kappa_n}$ and degree of freedom $2\alpha_n$. For further detail and derivations we refer the reader to Section 3 in [46].

The hyperparameter α controls the tails of the run length growth probability distribution. Larger values mean less mass in the tails of the distribution, with more data points being clustered around the mean. This means the likelihood of extreme run length values is small and changepoints are more prevalent. These results are antithetical to the MAP estimate derivations from Section 2.2. Meaning if the hyperparameters are poorly chosen, the resulting prior distribution may not accurately capture our beliefs about the true parameters, and this can lead to incorrect or biased posterior estimates.

As the value of α increases, Student's t-distribution becomes more spread out with more probability in the tails, resulting in less changepoints detected. β measures the deviation of the distribution from a symmetric distribution with zero mean. Similarly to α, smaller values of β means that the distribution is centered

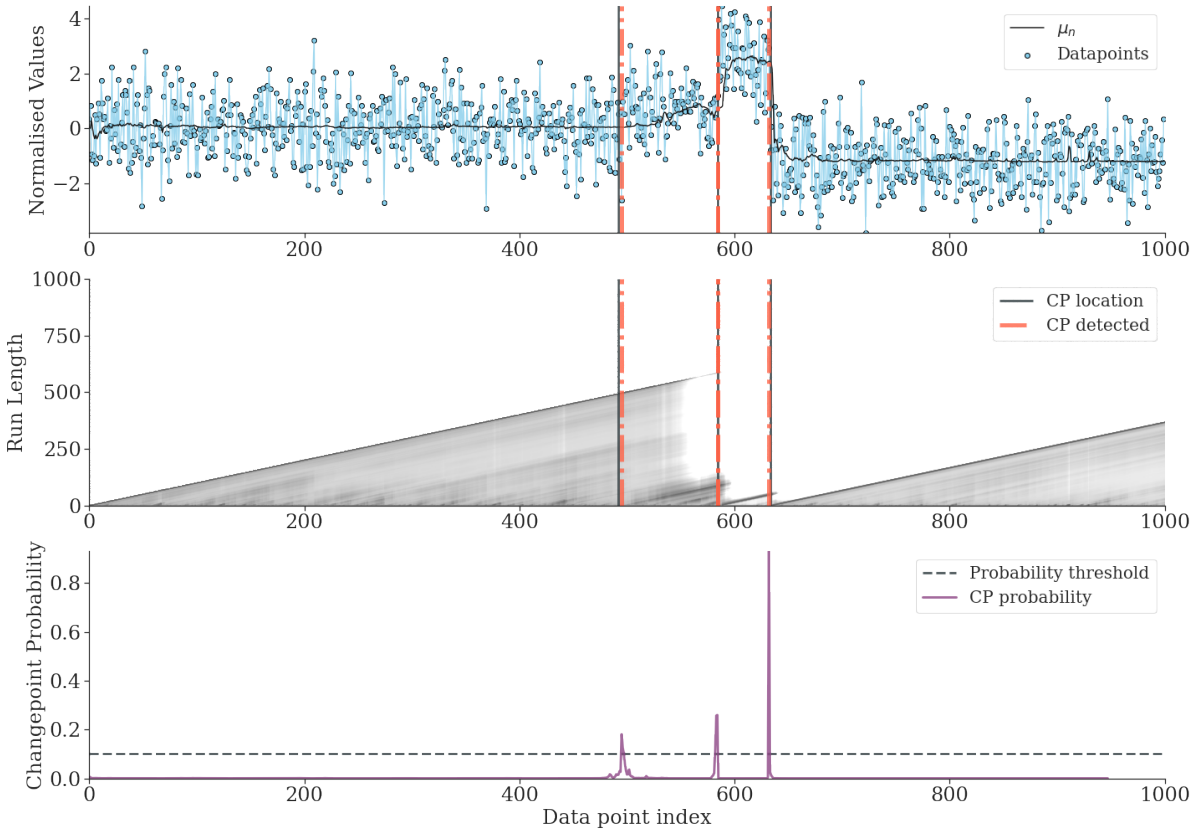


Figure 3.10: Bayesian online change point detection for $T = 1000$ synthetic data points modelled from a Student's t distribution. The growth probability of run length $\mathbb{P}(r_t, \mathbf{x}_{1:t})$ modelled from a Student's t distribution are also able to pick up the three change points with almost near perfection when presented with a large data set.

around zero. The Student's t-distribution becomes more skewed as the value of β increases. thus taking the truncated distribution, more extreme values for the run length growth are expected.

The hyperparameter κ refers to the precision of the estimates. This scale parameter is reciprocal of the standard deviation of the distribution. Thus as it increases the distribution becomes more spread out, resulting in fewer changepoints. In contrast, smaller values of κ means the distribution is more concentrated around the mean. Smaller possible values of run length growth mean more occuring changepoints. Finally, the hyperparameter μ represents the mean of the distribution. We further show the achieved results under various initial hyperparameter settings in Section 3.7.

Student's t-distribution looks almost identical to the standard Normal distribution but has more distribution in its tails. Both assume a normally distributed population but the probability of getting values very far from the mean is larger with a t-distribution. This has the benefit of mitigating the effect of outliers. The assumption here is that Student's t distribution should be used for modelling the early learning stages of using the Mako RAS system. With more observations the degree of freedom increases and the t-distribution approaches the standard Normal distribution. We show this method at work in Figure 3.10.

At each time step t, the predictive mean $\hat{\mu_t}$ is modelled from a Student's t distribution and is plotted using a solid black line. The predictive probability of the underlying model over the current run length π_t^r (Step 3 from Algorithm 3) is being modelled from a Student's t distribution with parameters $\alpha_0 = 1, \beta_0 = 0.1, \kappa_0 = 2, \mu_0 = 0$. The segment length and probability threshold are $l_t = 50$ and $p_t = 0.1$ respectively. The results of modelling the run length from a Student's t distribution in Figure 3.10 is akin to the results found in Figure 3.9. This is reasonable because as the sample size increases, the Student t distribution approaches Gaussian.

3.5.1 Multivariate change point detection

Akin to the offline method presented in Section 3.2.1, when modelling change point detection with Algorithm 3 for multivariate data from an online nature it is beneficial to model correlations between features using a covariance matrix. We opt to model the posterior predictive probability where the parameters in the multivariate case from Equation 3.22 are updated as in Wang et al. [36]:

$$\boldsymbol{\mu}_{n,p} = \frac{\kappa_{n-1} \cdot \boldsymbol{\mu}_{n-1} + \mathbf{x}_{1:n,p}}{\kappa_{n-1} + 1} \tag{3.33}$$

$$\kappa_n = \kappa_{n-1} + 1 \tag{3.34}$$

$$\alpha_n = \alpha_{n-1} + 0.5 \tag{3.35}$$

$$\boldsymbol{\beta}_{n,p} = \boldsymbol{\beta}_{n-1,p} + \frac{\kappa_{n-1} \cdot ((\mathbf{x}_{1:n,p} - \boldsymbol{\mu}_{n-1,p}) \cdot (\mathbf{x}_{1:n,p} - \boldsymbol{\mu}_{n-1,p})^{\mathrm{T}})}{2(\kappa_{n-1} + 1)} \tag{3.36}$$

Where the segments dimension is simply denoted using p. Figure 3.11 presents the multivariate BOCD method at work on change point detection for three Gaussian distributed data streams. The run length plotted as log scale probabilities in the middle plot clearly comes to an end with the start of a new partition segment. The assigned change point probabilities are also significantly high, although not as large as in Figure 3.5 when computing with Algorithm 2 which evaluates on all the data in a single batch due to its offline nature.

Multivariate BOCD applied to data with many partitions in Figure 3.12 is able to rely on the auxiliary variable in run length to assign higher probability of a change point occurring than in Figure 3.6. The benefit of this method is that we both take into account the multivariate covariance structure and obtain the same parameter update rule as in the univariate case without loss of generality [36].

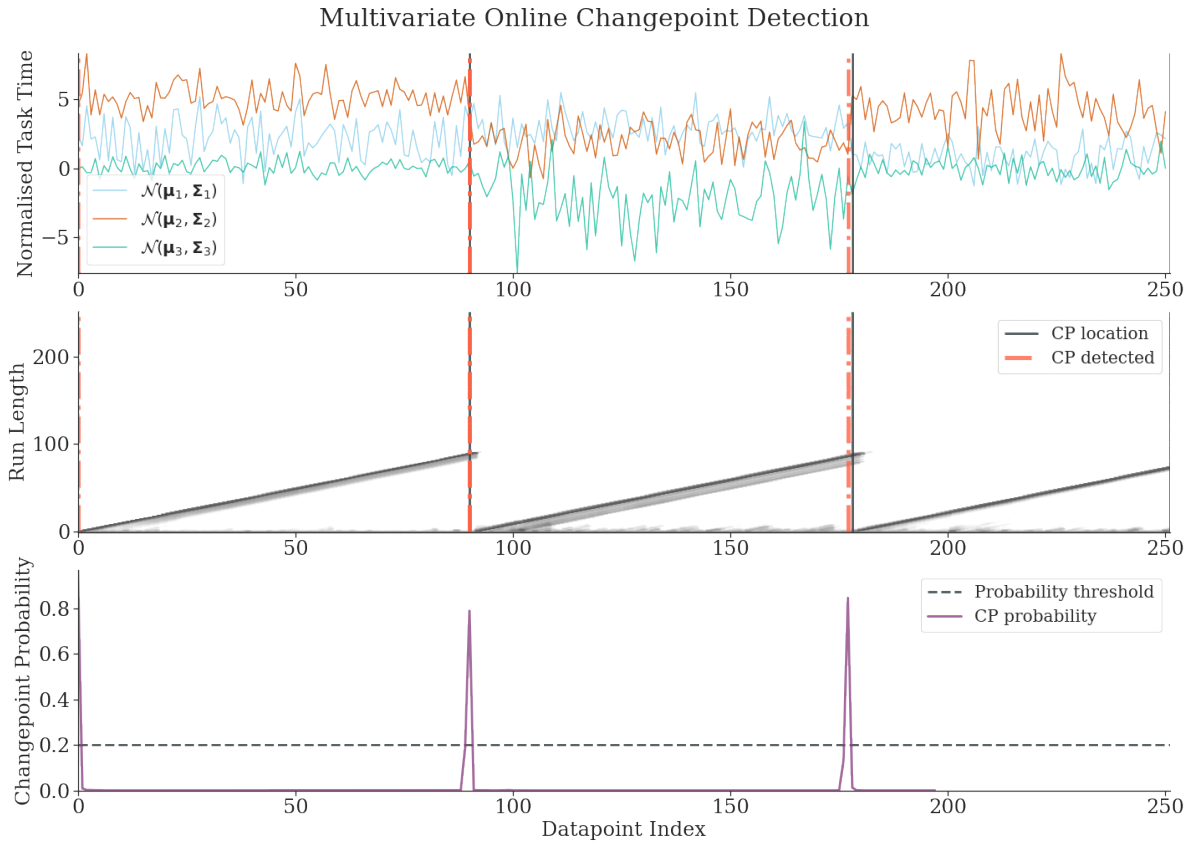


Figure 3.11: Multivariate BOCD for $T = 250$ synthetic data points modelled from three Gaussian distributions with change point locations at 90 and 178.

3.6 Small data set

The downfall of CUSUM analysis as was shown in Section 3.1.2 is that the number of observations heavily influences whether a learning curve is to be found, particularly when data is scarce such as when a new surgeon is employed by a clinic. BOCD can instead detect shifts in the data which is helpful because it remains invariant with respect to the amount of data used. We motivate further what happens when data is scarce. In particular examining whether Algorithm 3 is able to detect change points correctly over a shorter range of data points, and whether the growth probabilities on the run lengths should be modelled from a Gaussian distribution or a Student's *t*.

For this purpose we randomly generated $T = 30$ data points and again chose three different segments with varying mean. This index level was chosen because it is widely accepted that Student's *t* distribution outperforms the Gaussian distribution when observed data points total $n < 35$. The hazard function for a change point to occur at time t was set to $H(\cdot) = \frac{1}{15}$. The prior parameters for growth probabilities of run lengths modelled from a Gaussian distribution were set to $\mu_0 = 0, \sigma_0^2 = 0.5$. A larger value for known variance of the data $\sigma^2 = 2$ was chosen in order to highlight the increased variability in the performance of a surgical task we can expect to observe at the beginning of training. For Student's *t* distribution the prior parameters were chosen to be $\alpha_0 = 3, \beta_0 = 0.05, \kappa_0 = 0.1$. Since there are now less data points, the segment length was lowered to be $l_t = 5$ whilst the probability threshold remained at $p_t = 0.1$. The results of using the BOCD on this smaller data set are presented in Figure 3.14.

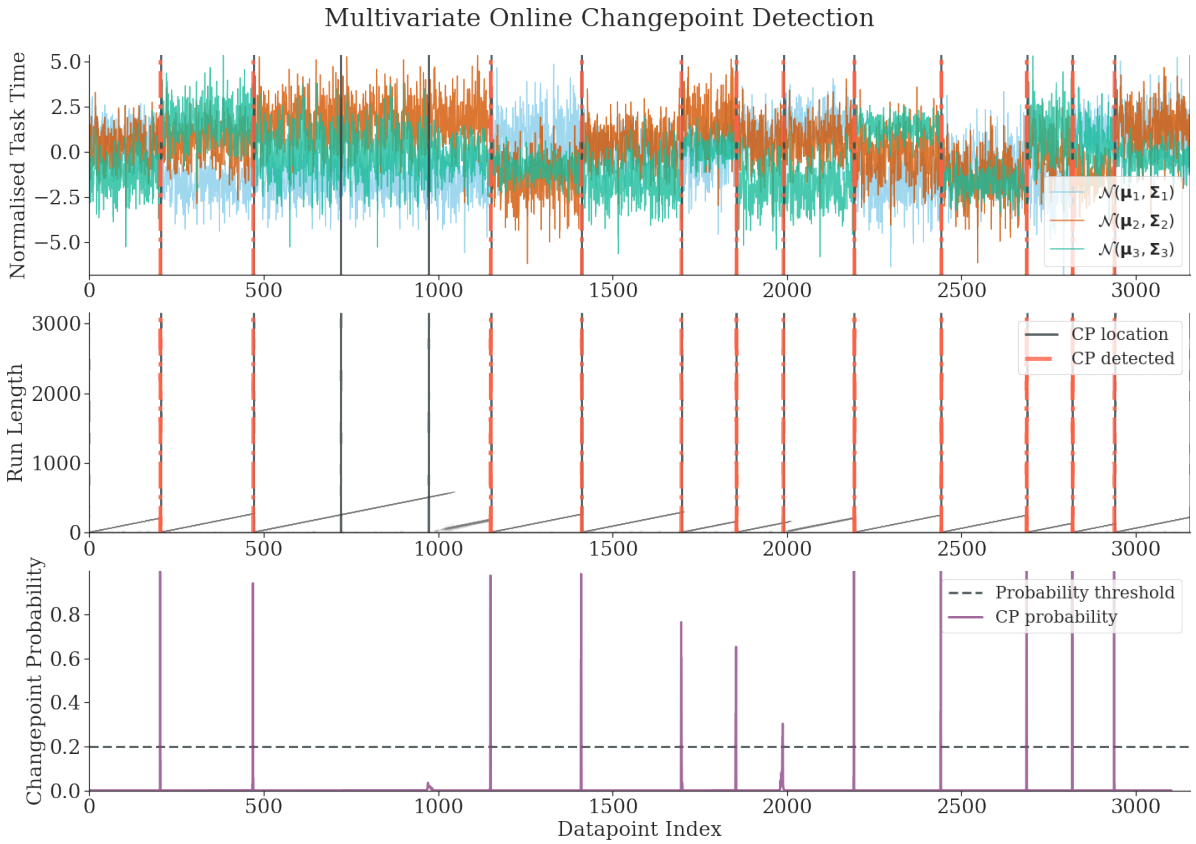


Figure 3.12: Multivariate BOCD for $T = 3156$ synthetic data points modelled from three Gaussian distributions with 15 change point locations.

3.7 Comparison between methods

Both the Offline BCD and BOCD algorithms are superior when comparing our results versus CUSUM analysis found in Figure 3.1. This is because the algorithms are better suited for detecting multiple shifts in the data process, as opposed to only one inflection point. This is more realistic therefore of a real world scenario where we might expect surgeons to follow various phases in their training such as from being an apprentice to intermediate, then to advanced, before finally becoming an expert in operating the MAKO RAS system. By bringing all our results together it is clearer why this is so.

3.7.1 Gaussian versus Student's *t* distribution

When it comes to modelling the posterior from a Gaussian and Student's *t* distribution then the results can be replicated between the former method with the correct choice of hyperparameters in the latter method. The top plot in Figure 3.13 and Figure 3.14 shows the normalised values over time, with the data points normalised around 0. The four different coloured vertical lines indicate real location of a change point, as well as where the three discussed methods identify a change point to have occurred. The middle plot tracks the run length probability. The bottom two plots compare using logarithmic colouring scale the probability of the current run length with growth probabilities being modelled from either a Gaussian or *t*-distribution.

Both BOCD methods are very good at detecting all three change points for large amount of data, achieving near perfect results. There is also no substantial difference between the bottom two plots in terms of run length distribution. On the other hand, using CUSUM analysis completely misses the mark of where either one of the three change points lie and incorrectly assigns the only change point it is able to detect.

The meager difference we observe is the run length modelled from a Student's *t* distribution over a

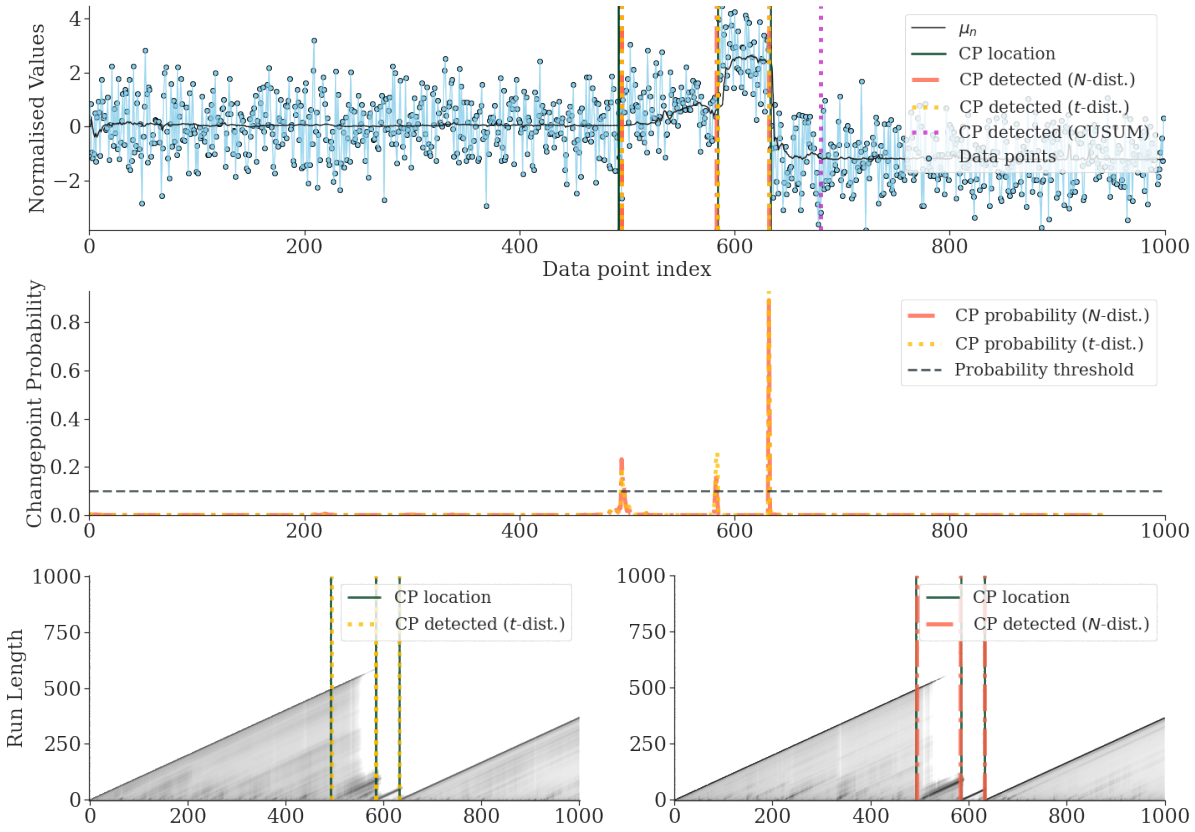


Figure 3.13: Comparison of change point detection methods for $T = 1000$ synthetic data points.

Gaussian distribution fitting the detected change point tighter to the real location of each change point, in particular the middle change point is identified exactly. Furthermore, Student's t distribution is able to compute the change point probability in accordance with the order of the change in magnitude of the varying mean between the four different segments. A larger shift in the mean between the segments is then given a greater change point probability in the middle plot. Despite observing little difference in the distributions in computing the posterior, we opt to model with Student's t distribution in Chapter 5 because we assume that we have no prior knowledge of the models parameters.

From Figure 3.14 we see that both BOCD methods perform well and detect the main change point at index 12. Student's t distribution outperforms the Gaussian distribution because it assigns a higher probability of change point occurring. It also picks up that a very early change point had taken place at index 1 and that the segment data that follows is different, something that the Gaussian is unable to distinguish correctly. Once again CUSUM analysis overshoots and assigns a change point prediction at index $t = 24$.

Furthermore, judging from the lack of variation in the pixel colour intensity inside of the bottom right plot, which depicts the growth probabilities of possible run lengths, the Gaussian assigns almost equal probabilities across the board. Albeit the BOCD identified a change point had taken place in Figure 3.14, the growth probabilities suggest that the BOCD algorithm, especially for the top diagonal, interprets most if not all of the data as belonging to the same run length. Concurrently, a different conclusion can be drawn from Figure 3.14 and made much more assertively. The bottom left plot clearly distinguishes the two different segments either side of the change point at index 12 as the shades of pixels distinctively contrast each other on the diagonal.

Lastly, the use multivariate models in Bayesian analysis are a complimentary addition when it is necessary to analyse several data streams. The posterior predictive distribution is able to pick up correlations between features and therefore identify a change point location more accurately. This has further benefit within a surgical workflow that involves completing many short subtasks. Employing Bayesian statistics paints a

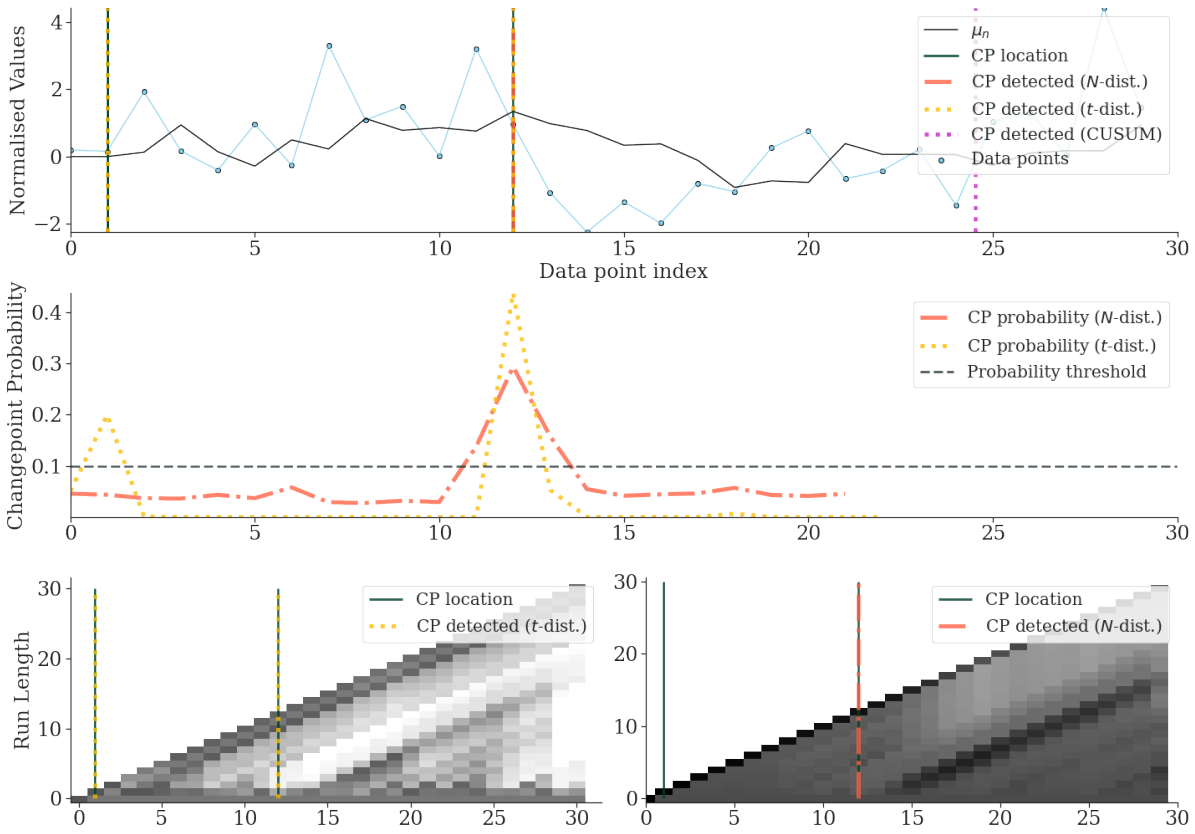


Figure 3.14: Comparison of change point detection methods for $T = 30$ synthetic data points.

fuller picture of the learning phases than with many univariate CUSUM analysis models and we ought to examine this on industry data in Chapter 6.

3.7.2 Offline versus online detection

In seeking to understand whether our analysis of surgical data would benefit from an offline algorithm when compared with one from an online nature, we witnessed that both perform equally well in the univariate case. What more is that both offline BCD and BOCD are able to leverage the covariance structure between data streams in the multivariate case. We run further experiments to see which method is superior under distinct values for the variance of the data.

One clear advantage the offline method holds is hindsight by iterating over the data twice using the EM approach. On the other hand, the online method can incrementally detect changepoints as the data arrives. This method also benefits from the initial hyperparameters that can determine where a change point is detected. Sholihat et al. [37] have extensively shown the role of parameters for efficient change point detection in BOCD.

The authors in [37] had shown using a probability distribution function for the predictive distribution of a newly observed data point since the last change point of a Student's t-distribution that larger initial hyperparameter values of α_0, β_0 increases the probability density function $\mathbb{P}(x_t|r_{t-1}, \mathbf{x}_{t-1}^r)$, which in turn decreases the probability of a change point occurring. Additionally, the possibility of a change point increases under small values of α_0, β_0 since decreasing the probability density function of the growth probability in turn increases the probability of a change point occurring. These results match the theoretical rationale presented in Section 3.5.

An additional hyperparameter λ is used to formulate the hazard function $H(\cdot) = \frac{1}{\lambda}$. Larger values of λ clearly decrease the hazard rate and with that also the probability of a change point decreases. On the

contrary, if λ decreases, then the changepoints inter-arrival time decreases, and the number of changepoints increases. For the synthetic data sets we keep λ unchanged at $\lambda = N$ for data sequence with length N. This was decided because under real life scenarios we often have no knowledge of how many changepoints exist. Therefore, we instead focus on the initial values of hyperparameters $\alpha_0, \beta_0, \kappa_0$.

To compare both offline and online methods we generated three synthetic data sets with length $N = 1117$ across which 15 changepoints occur. The segments of these time series are randomly assigned a length in the range of $50 - 80$ and a segment mean in the range of $0 - 1.5$. For the variance of each segment we selected from three possible range values of limited variance $0 - 0.1$, moderate $0 - 0.5$ and substantial $0.5 - 1.5$ in order to test the robustness of each algorithm against fuzzy data. Additionally, in order to put more emphasis on a change point occuring the variance of the data points increased by a factor of 0.01 across each segment before resetting at the next change point. With the Offline BCD a uniform distribution between two successive points $g(i) = \frac{1}{N}$ was used. For BOCD the hazard function used remained constant $H(\cdot) = \frac{1}{N}$ on the run length between two changepoints.

The experiments assess the precision accuracy of these methods and include three assessment criteria using a radius of δ on which to base the findings. These include the number of changepoints correctly identified within the range of $\delta = 5$, the number of changepoints correctly identified within the range of $\delta = 10$ and all the incorrectly identified changepoints outside of this radius. In Table 3.1 this precision accuracy is assessed on time series data with limited variance across segments.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	15	11	6	15
CP identified with $\delta = 10$	15	11	6	15
Incorrectly identified CP	0	0	0	0

Table 3.1: Precision accuracy comparison between Offline BCD versus BOCD under various hyperparameters. The variance of the data points was limited across all sixteen segments.

The Offline BCD correctly picks up all 15 changepoints, whilst incorrectly identifying zero others. BOCD is able to identify 11 changepoints with initial hyperparameters $\alpha_0 = \beta_0 = \kappa_0 = 1$, whilst the performance drops by identifying 6 changepoints when increasing the initial hyperparameters values. The BOCD is able to match the performance of the Offline BCD algorithm only when the initial hyperparameter values are lowered to $\alpha_0 = \beta_0 = \kappa_0 = 0.1$ with all 15 changepoints identified correctly.

Increasing the segments variance to moderate level brings about new challenges for both algorithms by having to contend with greater spikes in the time series that may or may not exhibit behaviour of a change point. The results in Table 3.2 show that the overall precision in terms of correctly identified change points stays as before and the Offline BCD continues to outperform the BOCD algorithm under various initial hyperparameter settings. Increasing the segments variance even further to substantial level in Table 3.3 brings about a decrease in the precision accuracy. The Offline BCD outperforms BOCD but also incorrectly identifies a greater number of changepoints that may had been caused due to greater spikes caused as a results of substantial variance inside the time series.

Intragroup analysis for the BOCD reveal that with smaller initial hyperparameter values the online algorithm is able to correctly identify more changepoints. These findings are inline with Sholihat et al. [37]. The stark difference is when we introduce substantial level of variance to the time series. In that instance, it is best to not choose the smallest initial hyperparameter values as shown in Table 3.3. Although this did come at a cost of incorrectly identifying more changepoints also.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	13	11	7	13
CP identified with $\delta = 10$	15	12	7	14
Incorrectly identified CP	0	0	0	0

Table 3.2: Precision accuracy comparison between Offline BCD versus BOCD under various hyperparameters. The variance of the data points was moderate across all sixteen segments.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	8	7	5	6
CP identified with $\delta = 10$	9	7	5	6
Incorrectly identified CP	3	2	0	1

Table 3.3: Precision accuracy comparison between Offline BCD versus BOCD under various hyperparameters. The variance of the data points was substantial across all sixteen segments.

3.7.3 Multivariate offline versus online detection

Similar experiments in the multivariate case are conducted in order to mimic the real world setting where several surgical steps have to be analysed concurrently. For this we generated three distinct time series of length $N = 1041$ using the aforementioned range values from the univariate case. The one common attribute of these time series is they all share the same change point locations. We begin with generating three time series with limited variance across segments in Table 3.4. We observe the Offline BCD to outperform the BOCD in terms of precision accuracy for identifying correctly the location of changepoints. Smaller initial hyperparameter values for the online method improve the algorithms performance in identifying all 15 change point locations within a radius of $\delta = 5$.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	15	13	2	15
CP identified with $\delta = 10$	15	13	2	15
Incorrectly identified CP	0	0	0	0

Table 3.4: Precision accuracy comparison between Multivariate Offline BCD versus Multivariate BOCD with various hyperparameters. The variance of the data points was limited across all sixteen segments.

Increasing the segments variance to moderate produces similar results in Table 3.5. The offline method continues to outperform the online algorithm despite tuning the initial hyperparameter to three distinct values. Increasing the segments variance to substantial we observe a significant drop in Table 3.6 in terms of precision accuracy when using the Offline BCD. Furthermore, the offline method also incorrectly identifies 2 change point locations that are caused due to fluctuations inside the segments. On the other hand, BOCD

with $\alpha_0 = \beta_0 = \kappa_0 = 1$ is able to identify 11 changepoints correctly across a radius of $\delta = 10$, whilst not mistakenly identifying any other locations as potential changepoints. It is therefore best for the practitioner to use the online method in the event that the multivariate time series data displays more variability between data points.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	15	13	9	15
CP identified with $\delta = 10$	15	13	9	15
Incorrectly identified CP	0	0	0	0

Table 3.5: Precision accuracy comparison between Multivariate Offline BCD versus Multivariate BOCD with various hyperparameters. The variance of the data points was moderate across all sixteen segments.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	6	10	7	3
CP identified with $\delta = 10$	7	11	8	3
Incorrectly identified CP	2	0	0	0

Table 3.6: Precision accuracy comparison between Multivariate Offline BCD versus Multivariate BOCD with various hyperparameters. The variance of the data points was substantial across all sixteen segments.

3.8 Summary

Throughout this chapter we had introduced CUSUM analysis, Offline BCD and BOCD algorithms, whilst exemplifying working examples and the robustness of each method in identifying change point locations. We had also ran experimental comparisons on the best practice of using these methods for small and large data sets, sampling from Gaussian or Student's t-distribution for the posterior probability, and tested for robustness of the offline versus online algorithms with univariate and multivariate data. In summary, this is how to use change point detection algorithms to the best of our ability:

1. CUSUM analysis is only able to detect at most one change point.
2. Offline BCD and BOCD algorithms allow for multiple change point detection through recursive updating of the run length probability.
3. Bayesian change point detection methods are more robust than CUSUM for detecting multiple shifts in univariate and multivariate data streams.
4. When the variance of data is low, offline method outperforms the online method. Conversely, both methods struggle under high fluctuations.
5. Online data processing can outperform offline batch data via the inclusion of an auxiliary probability variable evaluating whether the data belongs to the same run length.

6. Adapting the BOCD algorithm to the multivariate case is straightforward due to conjugacy in Bayesian statistics.
7. It is up to the user or practitioner to set the probability of a change point occurring, as well as the initial priors and hyperparameters.
8. In the online method, higher hazard function and smaller initial hyperparameter values decrease the run length growth probability and in turn increase the possibility of a change point.
9. For large data sets it is reasonable to model run lengths from a Gaussian distribution.
10. When data is scarce, run lengths modelled from a *t*-distribution are preferred due to being less prone to influence from outlier data points.

Chapter 4

Experimental results of learning curves

In this chapter we showcase and discuss results that help us on the way to answering Question 1. Each TKA surgery is comprised of several principal tasks that are recorded using the Mako RAS system. Some of these tasks originate in classical surgery, such as for instance the total surgery time and bone sawing time. Whilst other tasks are unique to the Mako RAS system. These tasks include but are not limited to the bone registration time, which is used to communicate with the system as to where the haptic boundaries are located, and ligament balancing, which assists the surgeon in setting anatomical alignment to match that of the pre-surgical plan in a patient. This makes the surgeons job more straightforward but also adds an additional phase in learning how to operate the new system.

To assess the learning curves in operating the MAKO robot, we take one task at a time and construct time series of the surgeons performance across consecutive surgeries. CUSUM analysis is then used in identifying whether inflection points on the learning curve exist and the surgeon is able to transition between the inexperienced to the proficient phase, thus displaying an elevation in the skill of using the RAS system.

We highlight how surgeons learn to perform RAS with the Mako by analysing at which point the transition from being the inexperienced to the proficient stage takes place and compare our work with alongside that of Vermue et al [32] , Kayani et al. [30] and Tay et al. [31]. This is done for the total surgery time from 299 consecutive surgeries, as well as the operative stage times of implant planning, ligament balancing, bone registration, bone cutting and bone sawing from 446 consecutive surgeries. Inflection points of learning curves are identified for all six tasks, however, the analysis had to be conducted across three time periods due to the Covid - 19 pandemic.

4.1 Total surgery time

The surgical time is described as the skin to skin contact time during which the surgeon is operating on the knee. This is from the time the first incision is made to when the last stitch is closed. We visualise the surgical time in minutes of surgeon 1 in Figure 4.1. Observe the gap in successive surgeries at the clinic of approximately three months between April and June of 2020 which occurred as a result of the world grappling with the Covid-19 pandemic. This world event forced the clinic to be shut until the safety of each workplace was assessed.

Once new legislation that prioritised the safety of all patients and staff could be implemented, the working environment was different. This is noteworthy of mentioning because from Figure 4.1 we observe a downward trend taking place leading up to the eve of Covid 19 in March 2020, showing that a surgeon is honing the skills in using the RAS system. However, immediately after the resumption of work in June of 2020, there is an evident change in trend of surgical time. This prolonged period out of work, combined with new health legislation being set for the workplace, resembles signs of a surgeon retraining in how to operate as part of this new and almost alien environment.

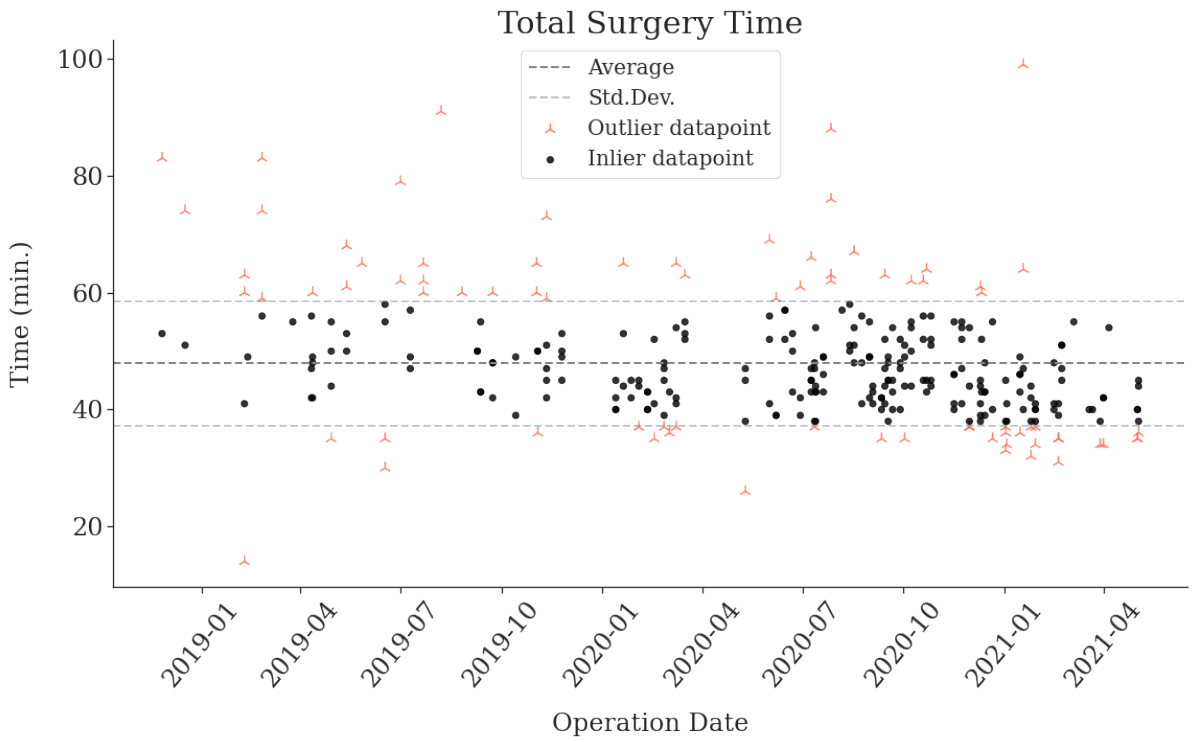


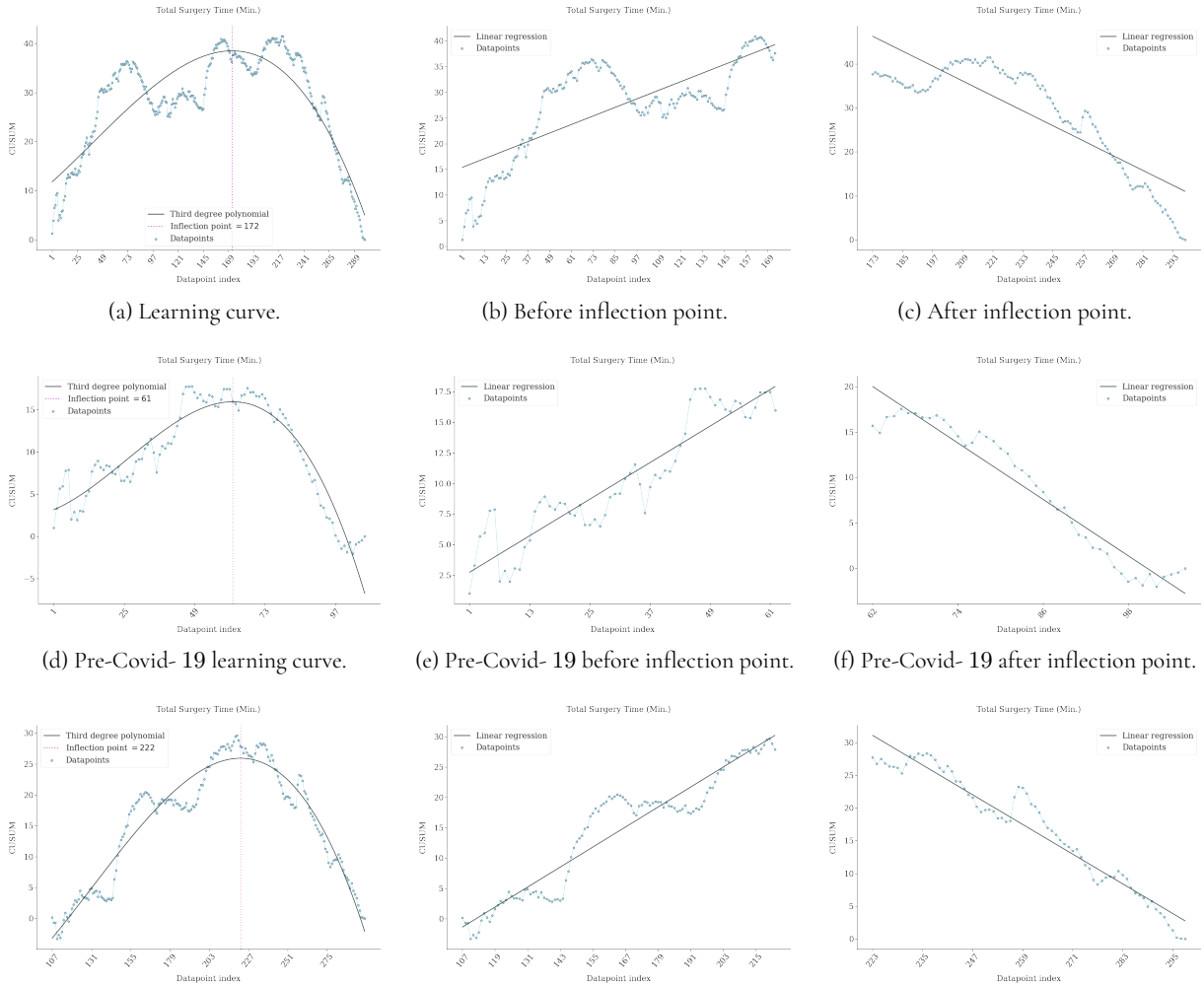
Figure 4.1: Surgeon 1 surgical time in minutes.

We therefore perform CUSUM analysis on three time periods in Figure 4.2. We do this across all TKA surgeries performed in the clinic by the surgeon, as well as splitting those surgeries into two periods: first period leading up to March 2020 we refer to as *pre-Covid*-19 and the second from June 2020 onwards we refer to as *post-Covid*-19. The latter period marks the period *post* when societal norms had changed and the clinic was forced to shut, as opposed to the complete eradication of the Covid 19 pandemic.

For the three aforementioned periods a third degree polynomial curve is fitted for each of the CUSUM charts and the inflection point is defined as being found at the global maxima. The phase before the inflection point is called the inexperienced phase and phase after is the proficient phase. Akin to Figure 1 in Kayani et al. [30], both phases benefit from being plotted to check whether the datapoints fit a linear regression trend. This results in a three by three grid of nine plots in Figure 4.2.

The number of surgeries included in the CUSUM analysis stood at 299, with 106 and 193 surgeries taking place pre- and post-Covid- 19 respectively. This total case load is threefold compared with the average case load of all three high case volume surgeons reported in Vermue et al. [32], fivefold greater than the cases reported in Kayani et al. [30] and tenfold greater than the average case load of the three surgeons reported in Tay [31]. An inflection point is revealed in Figure 4.2a after 172 consecutive surgeries performed by the surgeon. Our results show a much longer learning phase for the surgeon than only after 7 surgeries found in [30], the 11, 43 and 22 in [32] or the mean inflection point of 16 found in [31]. However, the maximum attainable CUSUM value only just surpasses 40, making it at a minimum four and at a maximum fifteenfold smaller than the values achieved elsewhere. There is little difference than than average attained CUSUM value in [31].

The visible issue for the learning curve found in our data is that the CUSUM chart exhibits two humped peaks either side of the date on which the surgery clinic closed for three months during the start of the pandemic. The first hump in Figure 4.2a at approximately 70 surgeries is followed by a sharp reduction in the running CUSUM value, indicating the surgeon began to improve with a reduction in the total surgery time before the pandemic. There is then an increase in the CUSUM value, which ultimately culminates in a second hump forming at approximately the 220 surgery mark, following the reopening of the clinic.



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

Figure 4.2: CUSUM analysis for the surgical time in minutes of Surgeon 1.

Fitting linear regression in Figure 4.2b and Figure 4.2c further exhibits that the CUSUM analysis method is not perfect when used in this scenario because the datapoints deviate further from the trend in both the inexperienced and proficient phases.

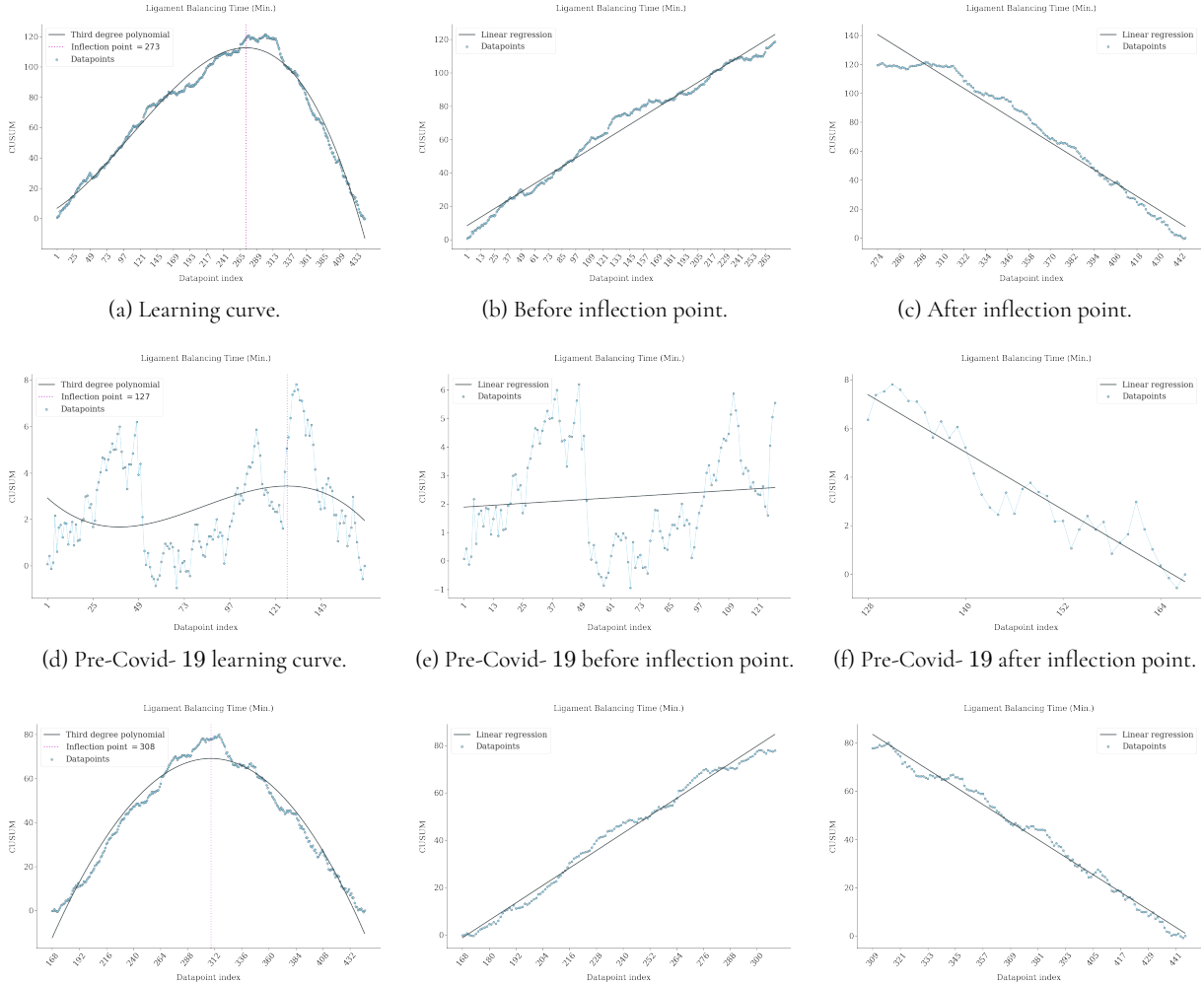
The learning curves for the pre- and post-Covid- 19 also exhibit much longer learning phases at 61 and 115 (222 if counted consecutively) surgeries. Once again the maximum attainable CUSUM values are much smaller when compared with [30, 32]. The linear regression exhibits a much tighter fit to the trend in Figures 4.2e, 4.2f, 4.2h, 4.2i. This further reaffirms our intuition that the CUSUM analysis method should be applied independently to each period, rather than bunched together for all consecutive surgeries before and after the pandemic.

4.2 Ligament balancing time

The precision of implant positioning utilises gap balancing, with the time to apply proper tension to the knee joint in extension and flexion being recorded via the ligament balancing time. The surgeon can then finalise the implant plan to obtain near equal medial and lateral gaps, as well as balanced extension and flexion gaps [38]. An assessment of the resulting joint gap balance is performed using the spacer block. The ideal knee gap balance generally has near equal joint tension in extension and flexion for medial and

lateral compartments. Asymmetric gaps may indicate that soft tissue releases or post-resection implant adjustments are necessary. From conversation with the three surgeons at the clinic it became clear that ligament balancing was a cumbersome task to learn when it came to using the Mako RAS system.

The number of surgeries used in the analysis for the ligament balancing time stood at 446, with 167 and 279 belonging to the pre- and post-Covid- 19 periods respectively. An inflection point of all the surgeries taken consecutively was found at 273 in Figure 4.3a. Both the inexperienced and proficient phases display good fits to the trend line in Figures 4.3b , 4.3c. With CUSUM analysis Vermue et al. [32] had found no learning curve for gap balancing. Both Kayani et al. [30] and Tay et al. [31] found a statistically significant difference for the ligament balancing time in minutes between the first ten cases and the rest, but with no learning curve being found for this operative stage.



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

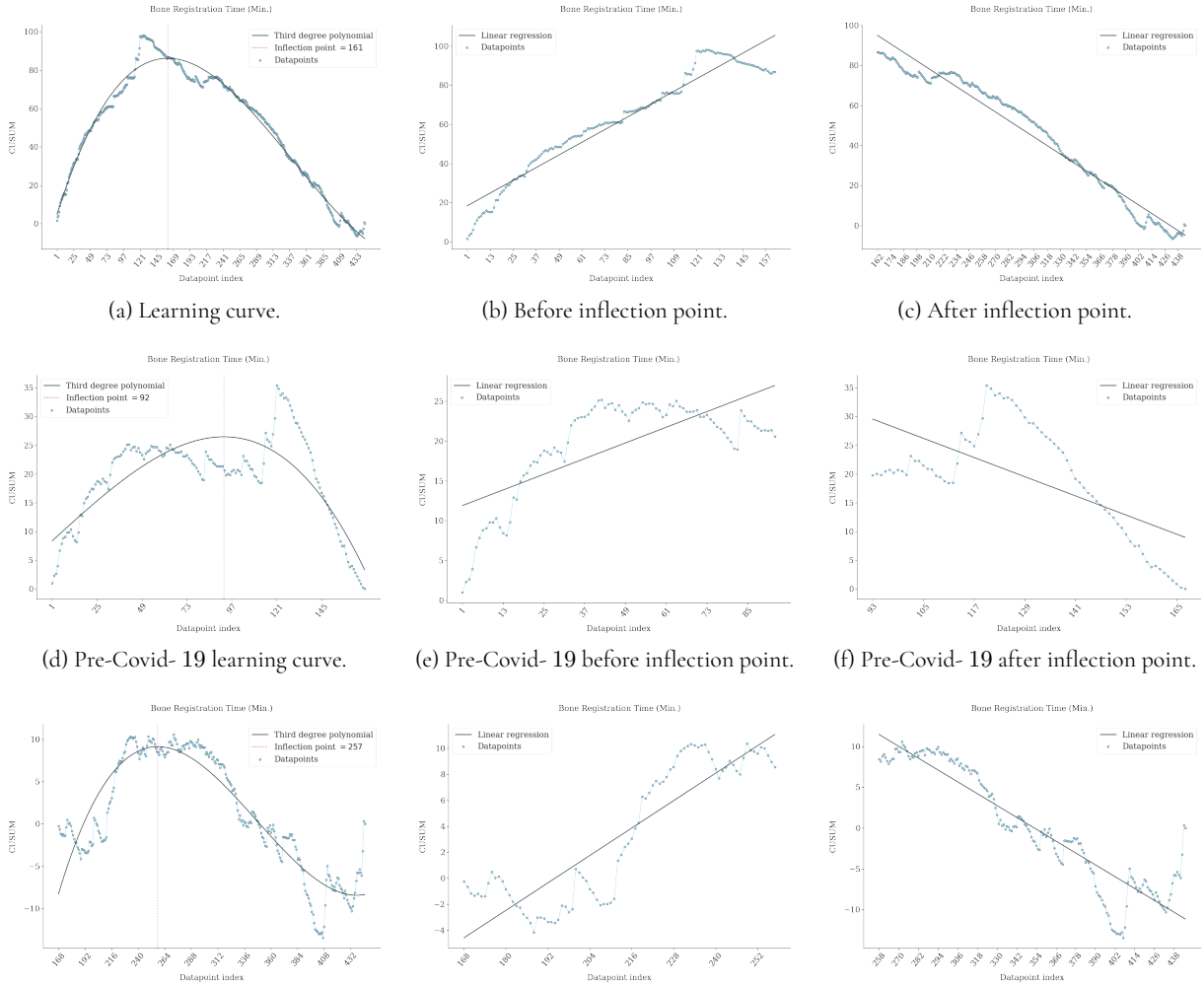
Figure 4.3: CUSUM analysis for the ligament balancing time in minutes of Surgeon 1.

When the pre- and post-Covid- 19 periods are analysed independently despite a global maxima point is found for the former after 127 surgeries in Figure 4.3d, this inflection point is not convincing due to its arrival after a global minima. This suggests that the surgeon worsened in terms of operative stage time rather than improved. As we know from the surgeons feedback, the complexity of this task may indeed require longer training time and hence we see more variability in the operative stage time. We conclude that no learning curve is to be found for the first batch of surgeries using CUSUM analysis. In the latter case a learning curve is found at 141 (308 consecutive surgery) in Figure 4.3g. Both phases fit the linear regression

trend very well in Figures 4.3h , 4.3i.

4.3 Bone registration time

Bone registration is a process whereby a surgeon collects point markers on the bony surface of the knee that in turn enables the RAS to track patient anatomy in real time. It is comprised of three distinct steps: patient landmarks (used in setting the mechanical axes of the bones), bone checkpoints (collects and verifies the checkpoints of the femur and tibia), and bone registration and verification (consisting of forty points the surgeon inserts a sharp probe tip into) [38]. To the non-surgical audience, this last step can be envisaged as looking at a screen whilst typing on a keyboard, a skill that does not come naturally at first but has the potential of improving with practice. Owing to the fiddly nature of these tasks, the bone registration is an extremely intricate step that allows for more precise implant positioning and resection during surgery. It therefore requires guile and experience to perform quickly.



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

Figure 4.4: CUSUM analysis for the bone registration time in minutes of Surgeon 1.

Akin to ligament balancing, the surgeons at the clinic had expressed similar notions of complexity in learning how to perform bone registration quickly and becoming proficient. We find an inflection point takes place at 161 surgeries in Figure 4.4a, with the inexperienced phase of the learning curve displaying a steep trajectory relative to the proficient phase thus indicating Surgeon 1 is able to improve fast in the short

term but slower in the long term. Both Kayani et al. [35] and Tay et al. [43] found statistically significant differences for the bone registration time in minutes between the first ten cases and the rest, but with no learning curve being found for this operative stage either.

Splitting the data into the pre-and post-Covid- 19 period we are still able to identify inflection points taking place at 92 in Figure 4.4d and 89 (257 consecutive surgery) in Figure 4.4g for both learning curves respectively. The linear regression trends fit poorly in Figure 4.4e , 4.4f , 4.4h , 4.4i. On the other hand, linear regression exhibits a tighter fit to the inexperienced and proficient trend lines in Figures 4.4b , 4.4c. This indicates to us that the real learning curve for bone registration is by taking all surgeries consecutively, with the pandemic having little to no impact on the ability of the surgeon in learning how to operate this surgical stage. Thus reaffirming the notion that Surgeon 1 improves faster when beginning to use the RAS system but this progress gradually slows across an extensive period of time.

4.4 Sawing time

During the bone sawing stage, the surgeon holds the saw blade and is guided along the haptic boundaries of the patients knee displayed on the RAS system screen. This is achieved with the help of the pre-surgical CT scan that makes up the 3 D virtual model of the patients knee. When the saw blade exits the haptic boundary zone, the stereotactic control of the saw blade is disabled, thus bringing no harm to the patient. When in cutting mode, if the saw blade exceeds 0.75 millimetres the boundary zone, the saw will not be powered. Longer bone sawing times can be indicative of the blade having to traverse a previously sawed area simply due to a lack in training.

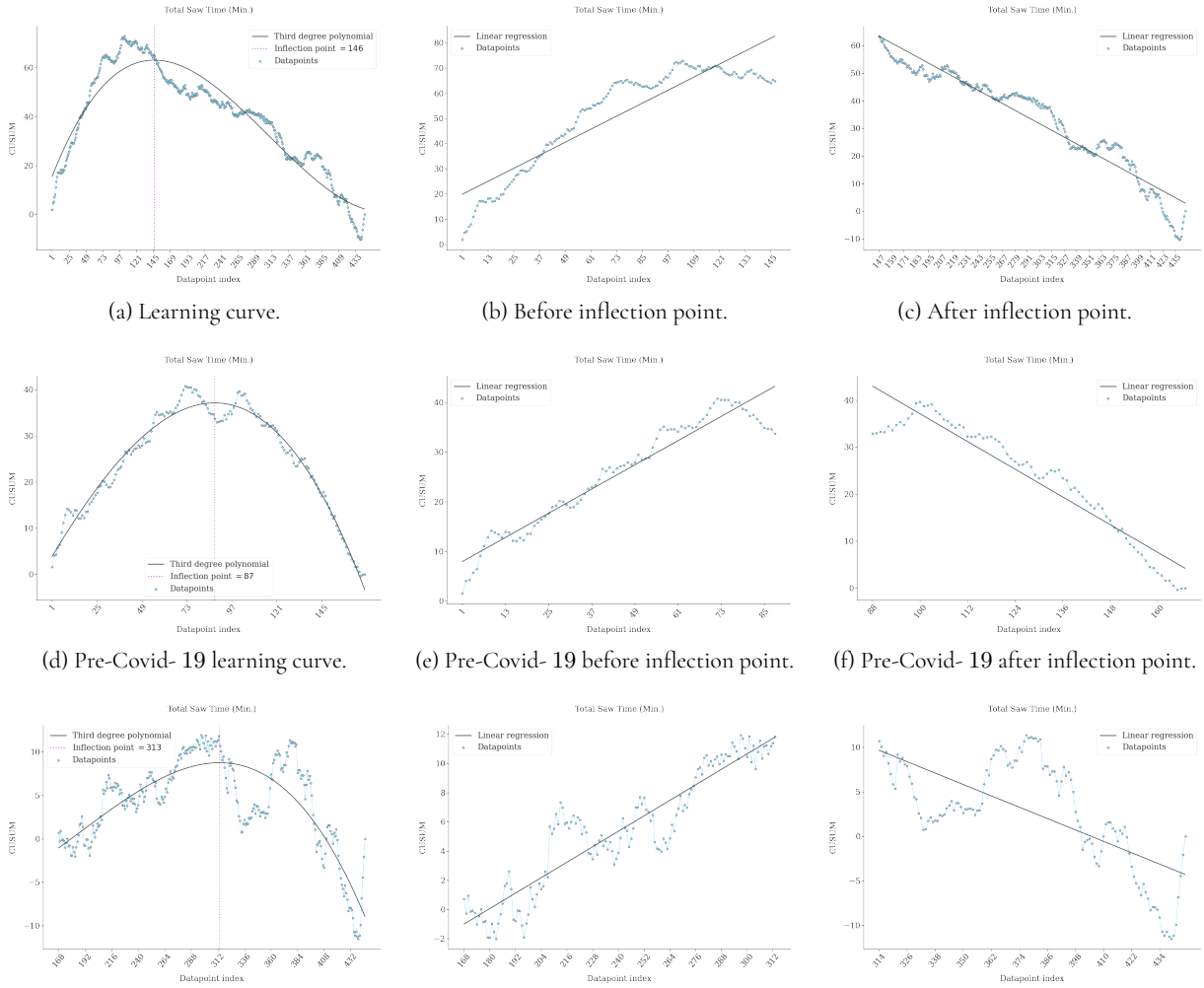
Out of the discussed surgical steps, bone sawing has the most resemblance between the surgical step in traditional analog surgery and with using the RAS system. The suspicion was that the learning curve for surgeons with prior experience of bone sawing would be flattened [48]. Instead, the learning curve displays a steep trajectory for the inexperienced phase and an inflection point is found at the 146 surgery mark in Figure 4.5a. This confirms the intuition that Surgeon 1 is able to quickly learn how to operate the bone sawing step by falling back on years of prior surgical experience.

Taking the initial 167 surgeries that took place in the pre-Covid - 19 period shows an inflection point for the learning curve at 87 in Figure 4.5d. Both linear regression trend lines fit the data well in Figures 4.5e , 4.5f. Although an inflection point is also found at 146 (313 consecutive surgery) for the post-Covid - 19 period, the learning curve in Figure 4.5c is less profound and attains a much smaller CUSUM value when compared with Figure 4.5d. This indicates that the time in performing bone sawing surgical step does not differ greatly over an extended period of time. We conclude the progress over an extensive period of time has slowed and it would therefore suffice to examine the learning curve of a surgeon using the CUSUM method for only the initial surgeries.

4.5 Summary

We have shown inflection points exist for the three surgical steps as well as the overall procedure. The results for both surgical steps implant planning and bone cutting time in Appendix B performed by Surgeon 1 further reaffirm that learning curves are identified when analysing Mako RAS data. In answering Question 1 regarding whether surgeons improve over many surgeries we can therefore give a concretely resounding *yes*.

The surgeon had demonstrated a long inexperienced phase for the ligament balancing and overall surgical time. These two tasks showed susceptibility to the events of the pandemic, exemplifying the difficulty in learning to use the RAS system. This shows that prolonged periods out of work for surgeons can result in longer operating times when using the RAS system. Both bone registration and bone sawing time showed relatively shorter inexperienced phases thus indicating the surgeon learned quickly. The pandemic had a lesser effect on the learning curve because the surgeon was able to improved faster at the beginning.



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

Figure 4.5: CUSUM analysis for the total saw time in minutes of Surgeon 1.

The data set we worked with was substantially larger which can explain why all the results showed relatively longer time until the inflection points and the transition between inexperienced to proficient phases than those found in [32, 30, 48]. Recall the caveats for interpretation from Section 3.1.2 on why this is the case when using the CUSUM analysis method. It is therefore not the best method to analyse data of many consecutive surgeries. Albeit its relative simplicity and intuitiveness, the variability in results shown between the two periods of pre- and post-Covid - 19 is also indicative of the flaws that the CUSUM analysis method provides. The surgical profession should therefore look elsewhere for better metrics in analysing and crucially comparing the performance between surgeons.

Chapter 5

Experimental results of surgeons improving

In this chapter we provide results that assist us in answering Question 2 and provide an alternative method that is able to build on the answer for Question 1. We do this by foregoing the CUSUM analysis method used in Chapter 4 and instead use the BOCD method. Recall that the threefold problem with the former method laid in the fact that it was incomparable across varying case number quantities thus comparison between surgeons becomes biased, more phases such as intermediate may exist other than only inexperienced or proficient, and finally the polynomial curve fitted during the CUSUM analysis may uncover a global minima instead which implies that surgical step time performed by the surgeon deteriorates.

Our results are directly compared with the findings from Chapter 4. Owing to the fact that all surgical steps differ in the nature of the task and hence in the time until execution, to perform the BOCD analysis we normalise the time in minutes of all the surgical steps and subsequently run Algorithm 3 over the time series. This was decided in order to retain consistency by keeping the initial hyperparameters α_0, β_0, κ_0 and μ_0 equivalent across all surgical steps. We alter the change point probability threshold between 0.1 and 0.4. Multiple learning phases are identified for all six tasks, displaying constant skills improvement by the surgeon. Furthermore, changepoints detected with BOCD analysis are invariant to different lengths of the time series across the three time periods as a result of the Covid - 19 pandemic.

5.1 Total surgery time

Once again we work with data total surgery time collected from 299 consecutive TKA surgeries. We find the BOCD allocates five independent phases to this time series, with change points identified in Figure 5.1 at 48, 148, 154 and 258. Because the middle two change points are near each other, we simply take the average of these two and assume we instead have four phases. Data for the total surgery time we denote as belonging to the novice $(1-47)$, intermediate $(48-150)$, proficient $(151-257)$ and expert $(258-299)$ phases.

The novice phase in Figure 5.1 clearly stands out from the other three phases due to possessing a larger mean value and greater variance of the surgeries. The initial period is therefore very important and can be achieved in under 50 surgeries, which is relatively faster than the 172 as suggested with CUSUM analysis in Figure 4.2a. The effect of learning diminishes in the subsequent two phases of intermediate and proficient as we observe only a slight reduction in both the average surgery time and variance. This slightest of differences is also observed in Figure 5.2b during our analysis of only the post-Covid - 19 data. The middle and bottom plots inside Figure 5.1 suggest that due to the low changepoint probability the BOCD also struggles to distinguish between the data across both phases. It is therefore possible that the data belongs to one longer phase from 48 to 257 surgeries.

The final phase is also clearly distinguishable from the rest with its low average value of surgical time and the variance of the data being more concentrated. The change point probability of a new phase beginning

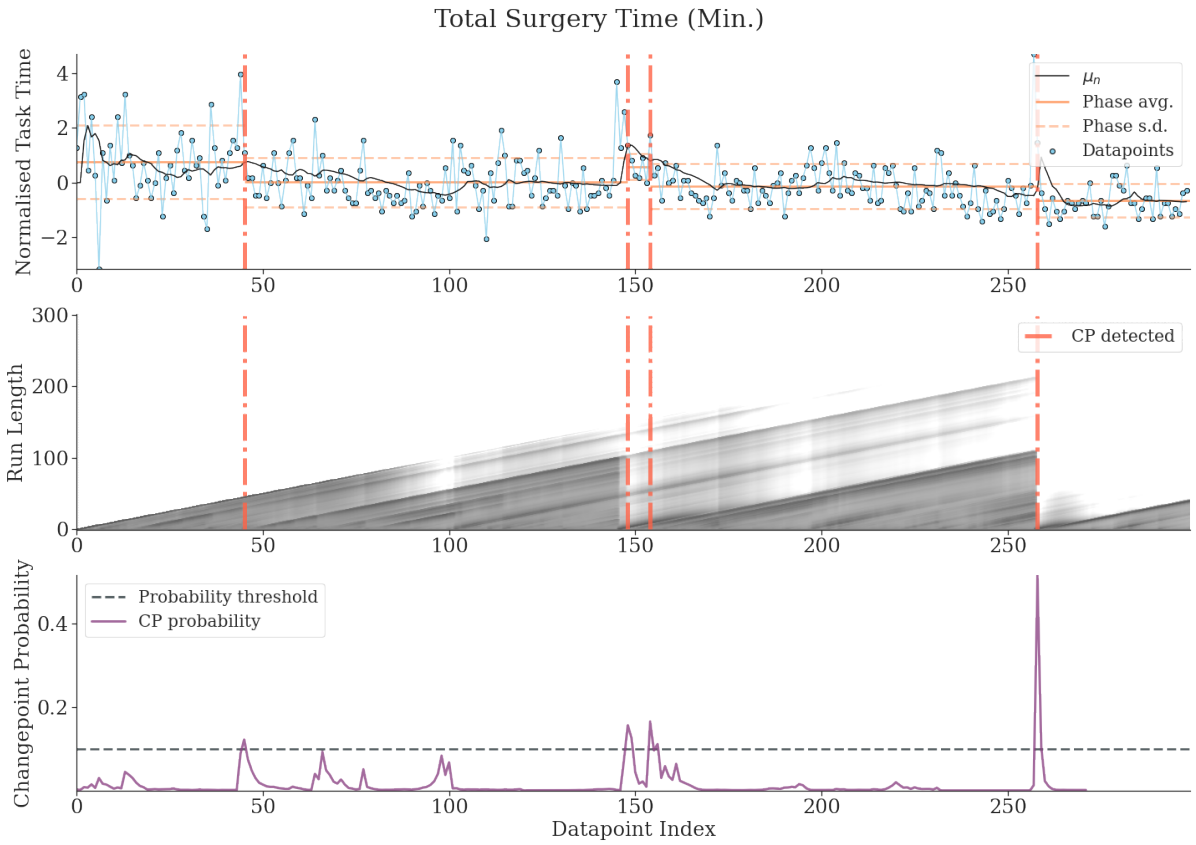


Figure 5.1: BOCD for the surgical time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

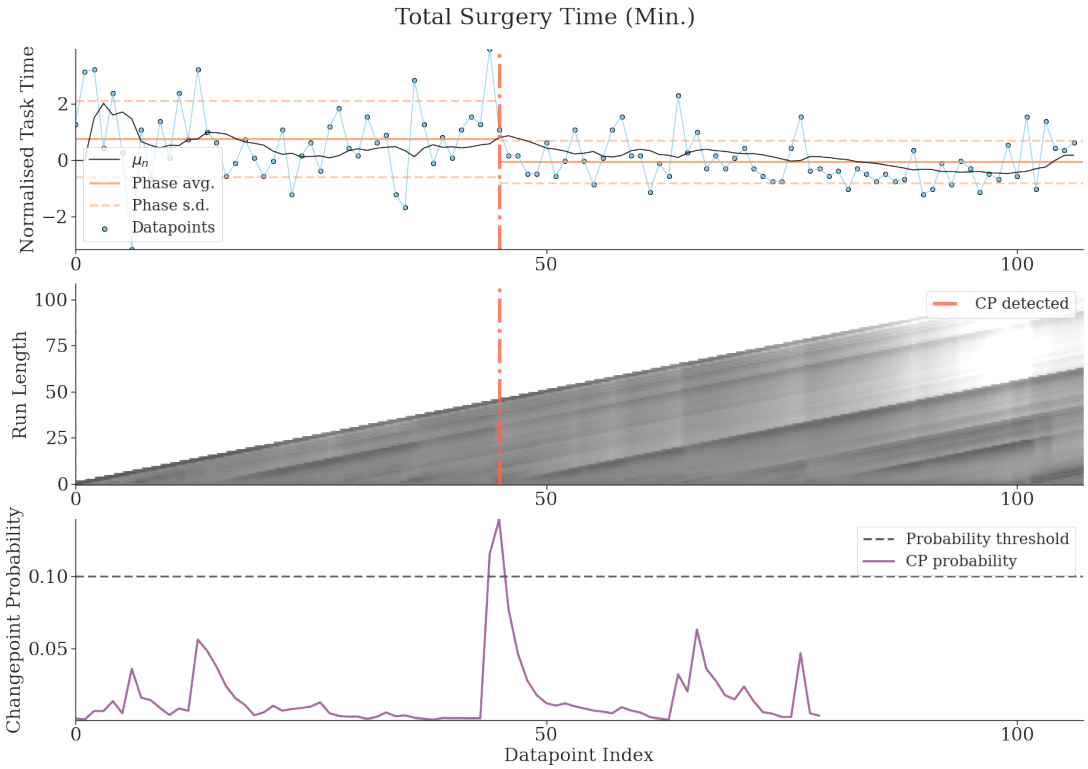
is also very high at nearly 50%. The surgeon becomes an expert in the final phase with the ability to operate quickly and almost indistinguishably in terms of time between consecutive surgeries. This suggests the surgeon is still able to improve much later in time than what was originally shown with CUSUM analysis.

When separating the data into the pre- and post-Covid - 19 periods the BOCD near perfectly identifies the same changepoints as before. Therefore we observe the overall ability of the algorithm to identify a change point is not determined by the quantity of data it has seen. The only difference being a change point is identified at 45 in Figure 5.2a rather than 48. The middle plot of Figure 5.2b shows that the BOCD algorithm continues to distinguish the fractional difference between the intermediate and proficient phases after a change point at 151 due to the diminished intensity of the pixels for the run length probability thereafter. The abnormally high surgical time for the two surgeries preceding this change point we believe contributes to a change point being placed at that location. Further work is necessary to explain the extent to which anomalous data points inside a time series impact on the performance of the BOCD algorithm.

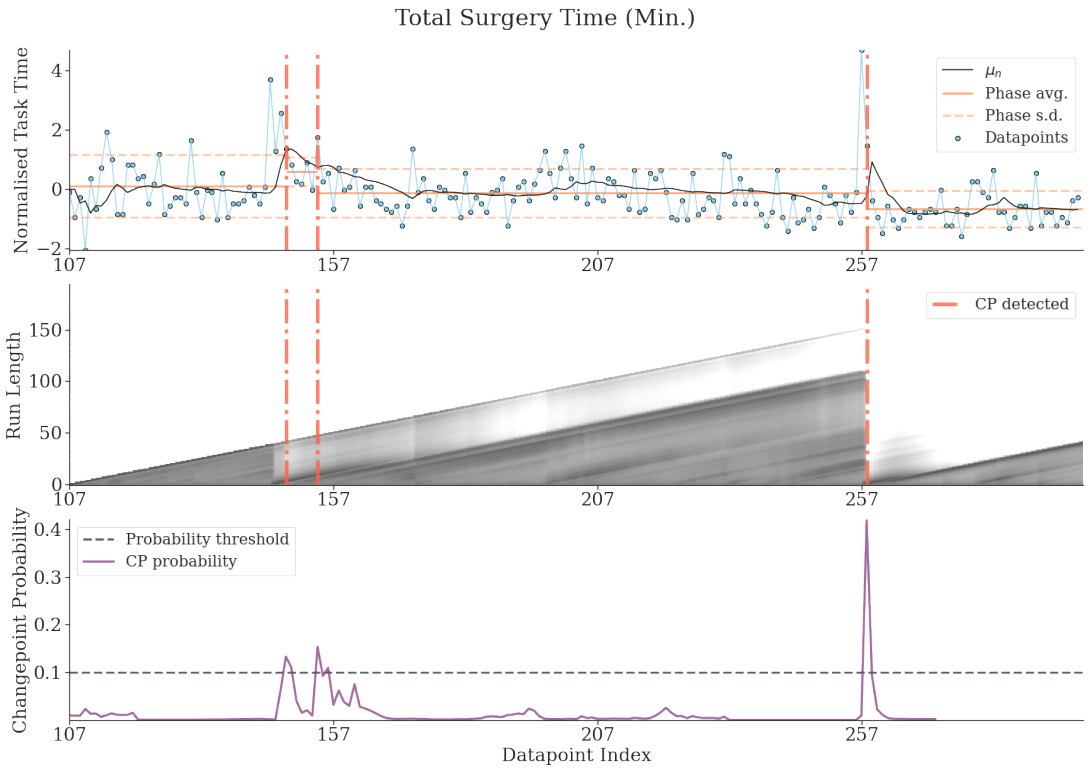
5.2 Ligament balancing time

The time series for the ligament balancing surgical step is modelled using 446 consecutive surgeries. The BOCD algorithm finds changepoints at 129, 265 and 316 in Figure 5.3. Hence the surgical learning phases occur during $1 - 128$ for the novice, $129 - 264$ intermediate, $265 - 315$ proficient and $316 - 446$ expert. Comparing with Figure 4.3a, the learning curve displays a somewhat similar inflection point at 273 until proficiency level is reached. We see that again it requires a long time until the surgeon reaches proficiency in performing ligament balancing together with the RAS system.

The stark difference in Figure 5.3 with Figure 4.3a is that we are better guided by the data in understanding



(a) BOCD for the surgical time in minutes of Surgeon 1 pre-Covid- 19.



(b) BOCD for the surgical time in minutes of Surgeon 1 post-Covid- 19.

Figure 5.2: BOCD for the surgical time in minutes of Surgeon 1 separated into pre- and post-Covid- 19 periods. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1$, $\beta_0 = 1$, $\kappa_0 = 1$, $\mu_0 = 0$.

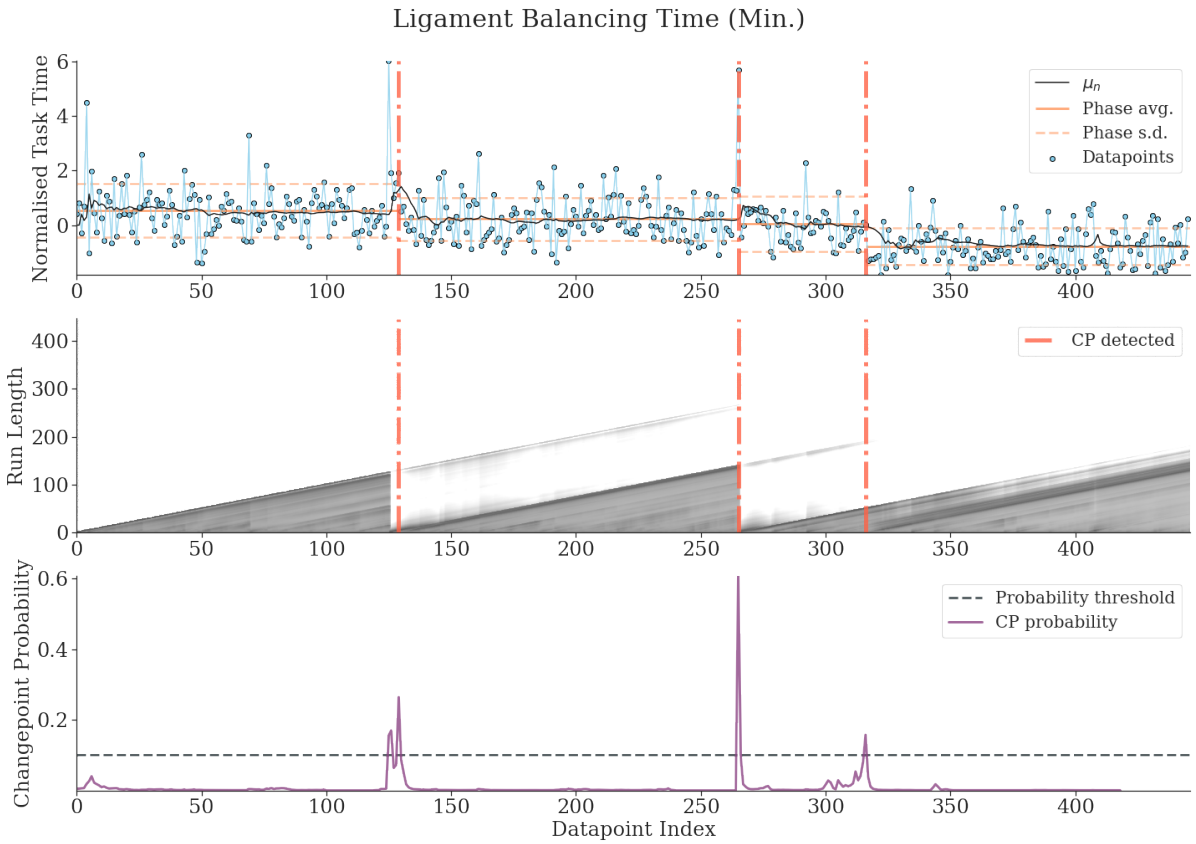


Figure 5.3: BOCD for the ligament balancing time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

how the learning takes place. Observe that the novice phase is marred with a greater mean ligament balancing time, as well as greater variability between consecutive surgeries, as opposed to the intermediate stage that immediately follows. We directly see how the surgeon improves with the ability to standardise the time in performing this surgical step. The final phase displays the lowest average ligament balancing time as well as the least variability between consecutive surgeries, therefore it is concluded the surgeon becomes an expert after performing 316 surgeries.

The changepoints in the pre - and post - Covid - 19 phases are identified at the exact same locations as given above and we therefore do not display these here. This again shows the BOCD algorithm is invariable to the amount of data it has seen. Note that the inflection point of 127 from Figure 4.3d is near identical to 129 in Figure 5.3. However, in the latter method we do not plot a learning curve and hence do not encounter a global minima existing before an inflection point. We are therefore not faced with the conflicting notion of a surgeon initially improving, then seeing a rise in ligament balancing time, before improving again for the final time as with CUSUM analysis. In another scenario it is possible that no learning curve in Figure 4.3d would had been found at all if only the first 100 surgeries were assessed. The BOCD also detects the proficient phase in Figure 5.3 much earlier than with CUSUM analysis in Figure 4.3g since there is a downward shift in the average ligament balancing time, albeit with higher variability between consecutive surgeries.

5.3 Bone registration time

For bone registration time we raised the threshold probability for the occurrence of a change point to 0.4 (or 40%) due to BOCD detecting many changepoints but assigning small probabilities to each. Changepoints are detected at 32, 120, 356, 371 and 411 in Figure 5.4. The novice phase between $1 - 31$ exhibits a larger

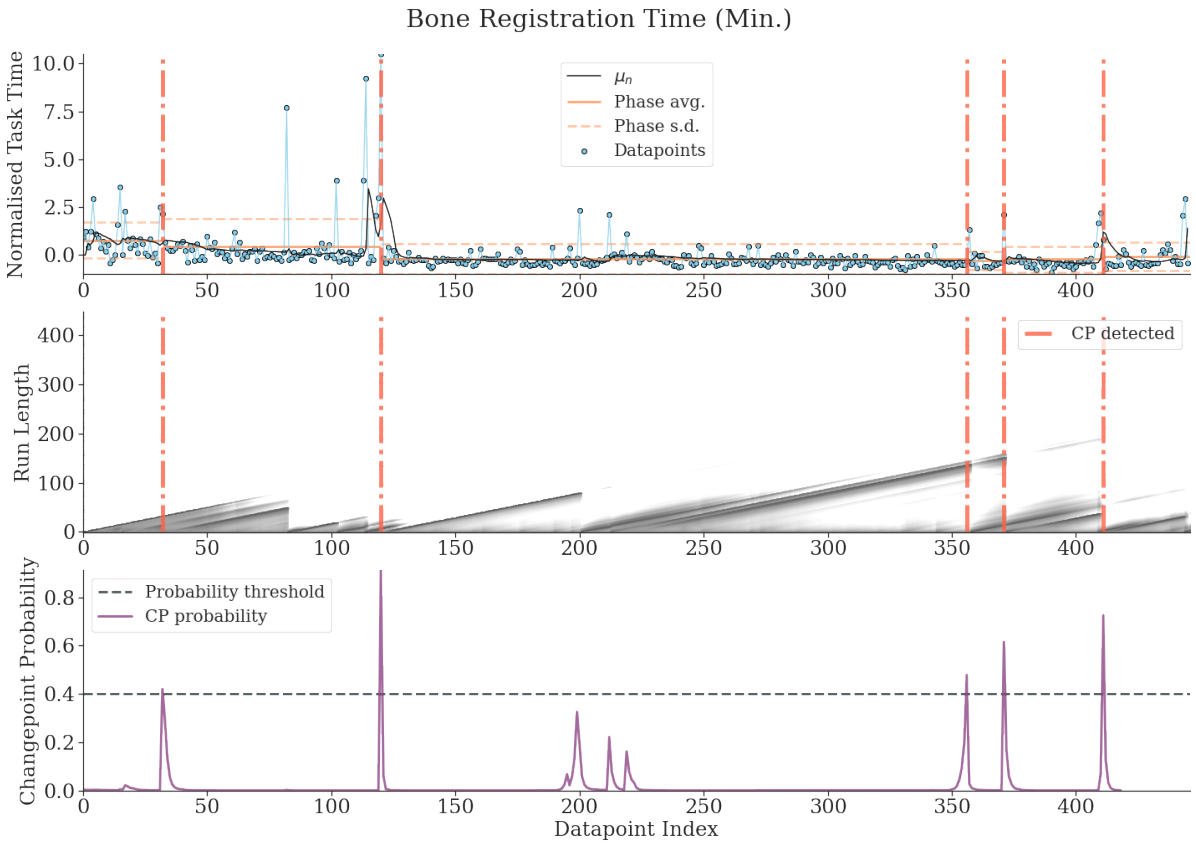


Figure 5.4: BOCD for the bone registration time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

average bone registration time than the subsequent phases. During the intermediate phase between $32-119$ the surgeon improves by demonstrating a lower average time to perform this surgical step, however, there is now more variability across consecutive surgeries.

The proficient stage $120 - 355$ proceeds for many surgeries and is reached relatively faster than with CUSUM analysis where an inflection point was found at 161 in Figure 4.4a. The BOCD method assigns a very high probability of over 90% to this change point, marking the start of the proficient phase at 120. Across both BOCD and CUSUM methods we spot a similarity with a short but steep period of learning followed by a prolonged period where little improvement is evident. This again suggests that the surgeon learns quickly in how to operate the Mako RAS system for the bone registration step, but across a prolonged period of surgeries this learning diminishes.

Immediately following the change point at 356 we observe a small increase in both the average phase time and standard deviation from the mean with the learning phase entering the final expert phase in Figure 5.4. The probabilities assigned to these change points are relatively smaller than when the surgeon entered into the proficient phase. It is therefore possible that these surgeries in fact belong to the same proficient phase and because we do not perform hyperparameter optimisation the BOCD algorithm assigns these higher than conventional bone registration times as another phase. Again we observe no difference in the changepoints locations and therefore do not show if the pre- and post - Covid - 19 periods are taken in isolation.

5.4 Sawing time

BOCD applied to the total sawing time detects changepoints at 8, 208 and 322 in Figure 5.5 when threshold probability was set to 0.4. This surgical step is of course a familiar application for a surgeon

trained in conventional surgery. The novice phase is thus very short with only seven surgeries necessary for the surgeon to familiarise oneself with removing the worn and damaged area of the bone. The surgeon improves little through the intermediate, proficient and expert stages with all three phases displaying near identical average phase time.

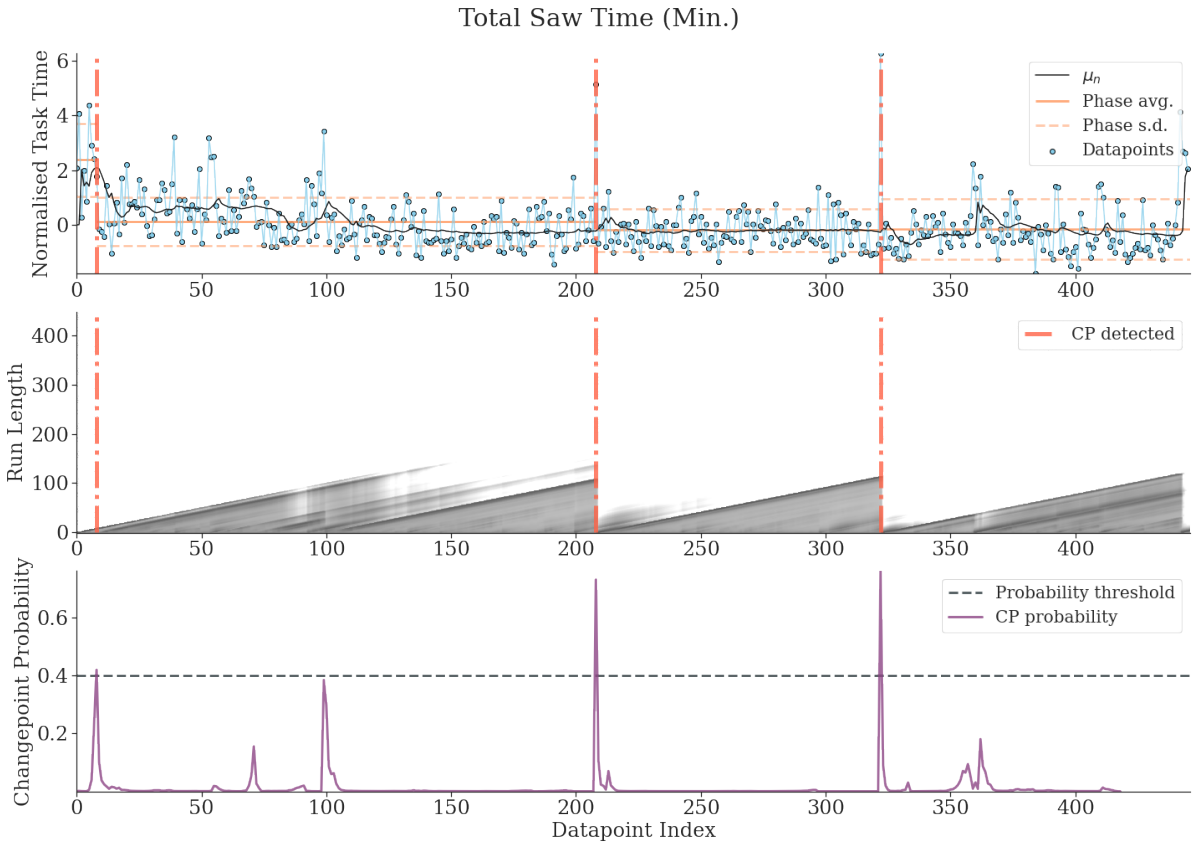
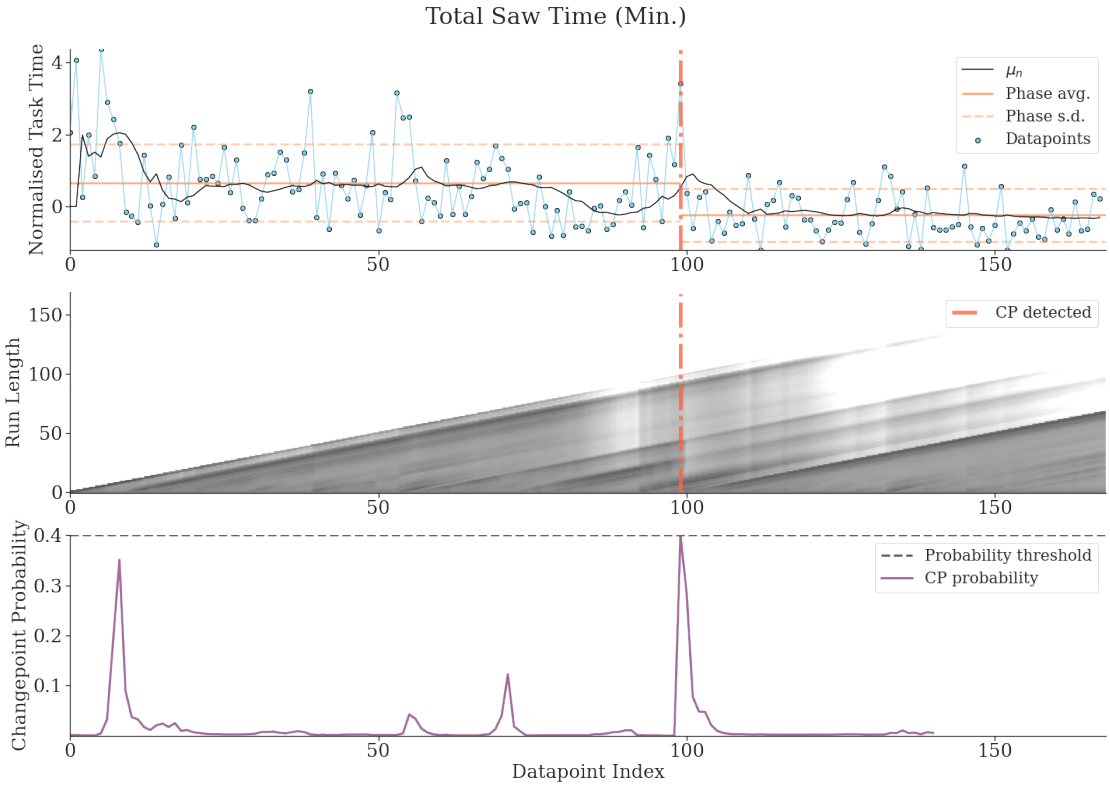


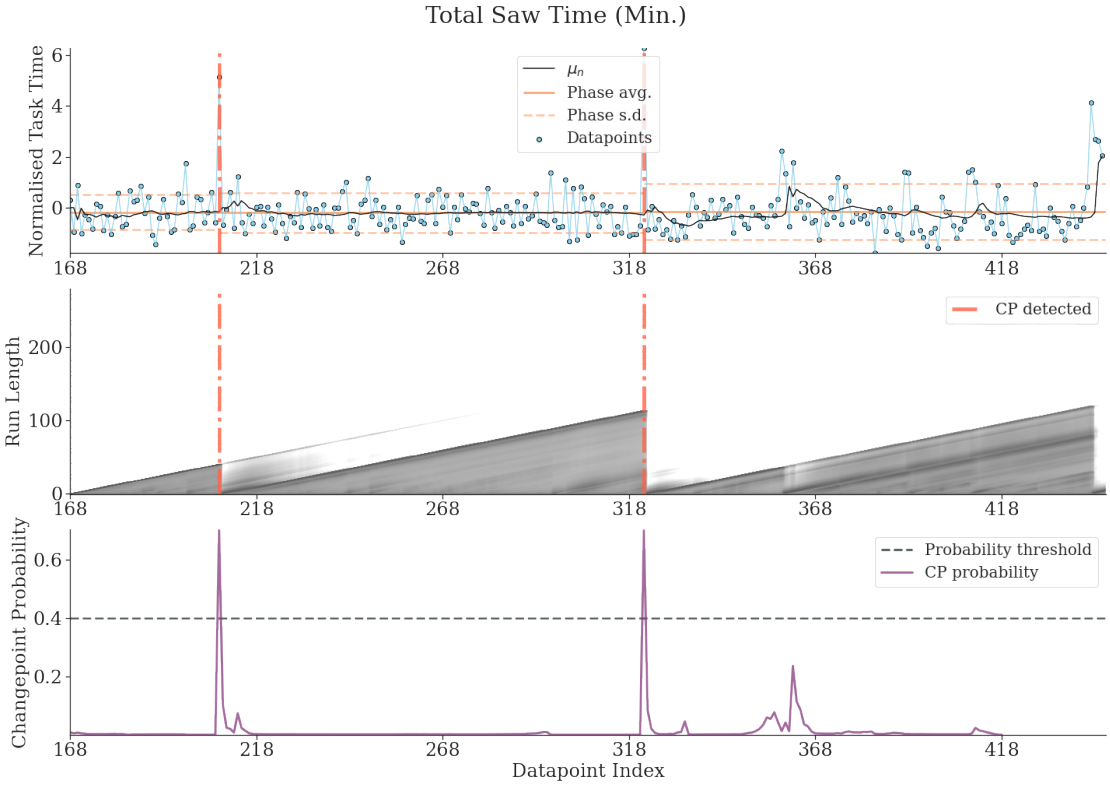
Figure 5.5: BOCD for the total sawing time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

However, as a result of taking the pre-Covid - 19 data in isolation the probability of a change point happening at 8 is now lower in Figure 5.6a and instead a change point is only detected at 99. Setting a higher threshold probability for a change point therefore returns a much longer novice phase. Looking at the characteristics of data the two phases clearly differ, with the latter phase in Figure 5.6a displaying lower average phase time with variance. In either case, the longest phase is the intermediate, hence the BOCD method detects a proficient phase starting at 208 in Figures 5.5, 5.6b later than the inflection point given with CUSUM at 146 in Figure 4.5a.

The results of using BOCD for the analysis of learning phases in total bone sawing time across the entire data array are not convincing because the detection of a change point is highly influenced by several abnormally large surgical phase times being recorded. What more, the data does not account for the burr size used. Smaller blades are more precise but are less powerful and thus the recorded surgical stage time can be higher. What is telling however is the algorithm is better suited than CUSUM to analyse tasks that a surgeon may be somewhat familiar already. The data points to BOCD relating quicker to the clinic that the surgeon is improving from the onset of using the Mako RAS system, helped by the use of the Student t prior for modelling the run length distribution.



(a) BOCD for the total sawing time in minutes of Surgeon 1 pre-Covid- 19.



(b) BOCD for the total sawing time in minutes of Surgeon 1 post-Covid- 19.

Figure 5.6: BOCD for the total sawing time in minutes of Surgeon 1 separated into pre- and post-Covid- 19 periods. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1$, $\beta_0 = 1$, $\kappa_0 = 1$, $\mu_0 = 0$.

5.5 Summary

We have successfully shown with BOCD analysis that for each of the surgical steps there exist four learning phases the surgeon traverses through in order to hone the skills in using the Mako RAS system. In the process we had exposed that using CUSUM analysis to identify a singular inflection point is insufficient and therefore discounts the true stages of learning an arthroplasty surgeon exhibits in practice. Particularly, the surgeon exhibited continued improvement for the total surgery time, ligament balancing time and bone sawing time notwithstanding having already completed over 300 surgeries. BOCD applied to the bone registration data showed inconclusive evidence caused by the algorithm applied to this time series being highly influenced by abnormally large recorded surgical step times. In answering Question 2 our answer is then *with BOCD we can uncover multiple learning phases at various locations for each of the surgical steps*.

Furthermore, we had again uncovered learning phases for the ligament balancing time, building on from the earlier work of Kayani et al. [30] and Tay et al. [31]. The BOCD analysis method is able to achieve more because it searches for transition between phases over time of the underlying process. It can thus better inform the clinic when the surgeons performance both improves and deteriorates. Whereas CUSUM analysis only communicates a singular change point in the process. The use cases with BOCD extend to when the surgeon if off work for an extended period of time such as during a global pandemic or illness. We are also better placed to provide meaning as to why a change point occurs by assigning probabilistic weight to how much the data in a particular phase differs from the rest of the time series as with BOCD. Extending our answer to Question 1 we also add that *BOCD routinely detects improvement by the surgeon inline with the assumption of Student t distribution for the univariate case*.

Chapter 6

Experimental results of golden standard

In this chapter we provide results that help with answering Question 3 and on the way discuss why arthoplasty surgery consisting of a handful of intricate surgical steps needs to be taken as a whole to correctly assess the golden standard in RAS. For this we forego analysing the total surgery time. The reasoning is twofold: the data gathered of total surgical times belongs to a shorter time series than the 446 data points of all other surgical steps, whilst also wanting to test whether BOCD is good at picking up on correlations in the learning phases between the surgical subtasks. For this we experiment with all five surgical steps together, implant planning with ligament balancing time, ligament balancing with bone registration time and bone cutting with bone sawing time.

The results with CUSUM analysis method from Chapter 4, as well as the work presented in Kayani et

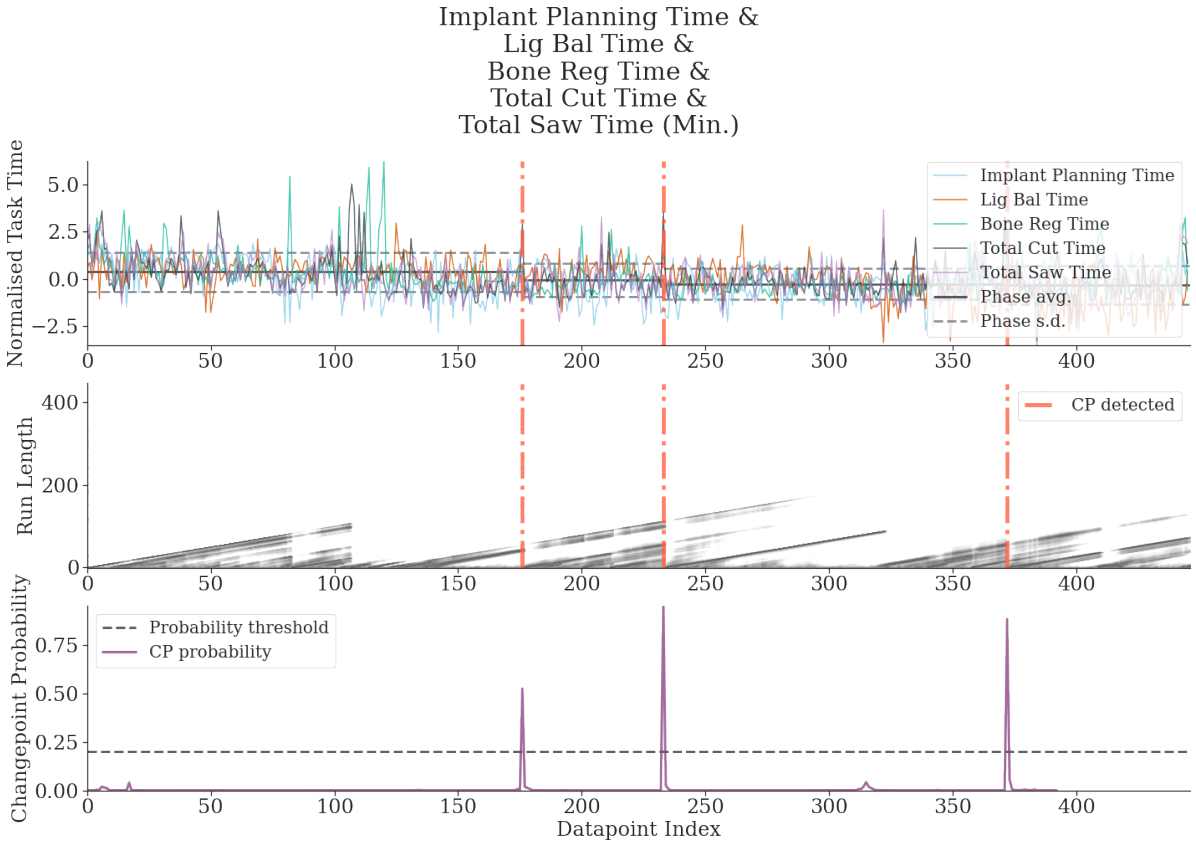


Figure 6.1: Multivariate BOCD for five individual stages of the robotic procedure in minutes of Surgeon 1 post-Covid- 19. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

al. [30] and Tay et al. [31], had only examined surgical steps as mutually exclusive events. In Chapter 4 and Chapter 5 we had examined the total surgery time as a metric for a surgeon improving. This is a good starting point for an arthroplasty clinic to begin assessing their surgeons. It was then shown how various learning phases exist for a multitude of surgical steps which helps towards understanding where more training is required. However, this only gives one perspective. The foreseeable issues that the clinic will otherwise run into can for instance be an increase in workload with having to monitor a multitude of surgical steps and difficulty in explaining whether the change in performance of one surgical step has an effect on another.

In a real world setting subtasks are often dependant on one another, such as the surgeons ability to prepare a good preoperative plan to help with sizing, aligning and positioning of the implant is important when intraoperatively fitting the prosthesis to the bony anatomy later on. What our analysis was therefore missing and which can be resolved with the help of data science is firstly a multivariate method to summarise all the tasks inside a single metric and secondly a more robust method that allows for a covariance matrix between features to be incorporated.

That is why we examine all five surgical steps from earlier using multivariate BOCD in Figure 6.1. The algorithm detects a change point between phases at 176, 233 and 372. Here the novice phase is the longest, spanning $1 - 175$. It is interesting to notice that the middle plot in Figure 6.1 displays run length ending after approximately 100 surgeries, however, the change point probability remains near zero in the bottom plot. Meaning the algorithm assigns low probability for the data from the first few surgeries belonging to a run length beyond this interval, being most likely caused by the cessation in fluctuation of the surgical steps time. Once the run length probability resets back to zero, the algorithm does not differentiate the surgeries that follow from the ones preceding the large fluctuations in the series. It requires more surgeries for the algorithm to gather evidence of a new phase beginning, leading to a change point only assigned at 176.

The intermediate phase in Figure 6.1 continues for $176 - 232$, displaying both a lower average phase time and smaller variance than during the novice phase. The surgeon reaches proficiency at 233 and again shows improvement over the previous two phases. An expert phase is reached in 372 surgeries and displays an almost identical average phase time as during the proficient phase only with a slight increase in the standard deviation from the mean, most likely caused with an increase in bone registration time as we had previously observed in Figure 5.4. Multivariate BOCD is useful in this context by providing a single metric to be used across an entire surgical procedure consisting of many individual yet correlated subtasks. It is therefore possible for the clinic to *mix and match* the surgical steps in order to create customised monitoring plans of each surgeon depending on their prior experience.

As previously mentioned, a well worked pre-surgical plan will allow for a more accurate implant alignment intraoperatively. Accurate alignment intraoperatively translates to speedier ligament balancing time. Both the implant planning and ligament balancing surgical steps had therefore been analysed together with multivariate BOCD in order to assess how the learning phases of both tasks impact on our assessment of a surgeons proficiency. Recall the univariate BOCD algorithm detected changepoints at 129, 265 and 316 in Figure 5.3 and at 100, 112, 360 in Figure B.3.

The multivariate algorithm detects a change point between phases at 74 and 344 in Figure 6.2. Taken together therefore, the novice phase for the two surgical steps is much shorter at only $1 - 73$ than for each task independently of each other. We observe the surgeon improves after a change point at 74 due to a distinctly lower average phase time during the intermediate phase. To reach proficiency or expert level it takes the surgeon until 344. Again this phase displays a lower average phase time than the previous two phases, communicating again that the surgeon is able to further improve. The number of surgeries required to reach the final learning phase is relatively similar to what we previously observed in Figure 5.3 and Figure B.3.

The distinct advantage over the univariate case can be seen in the final phase. Observe how the implant planning time is on the rise. If analysed with CUSUM we obtain a global minima point in Figure B.1a and an indication of deterioration rather than improvement with BOCD in Figure B.3. However, an increase in longer implant planning time can be translated into shorter ligament balancing time. The multivariate

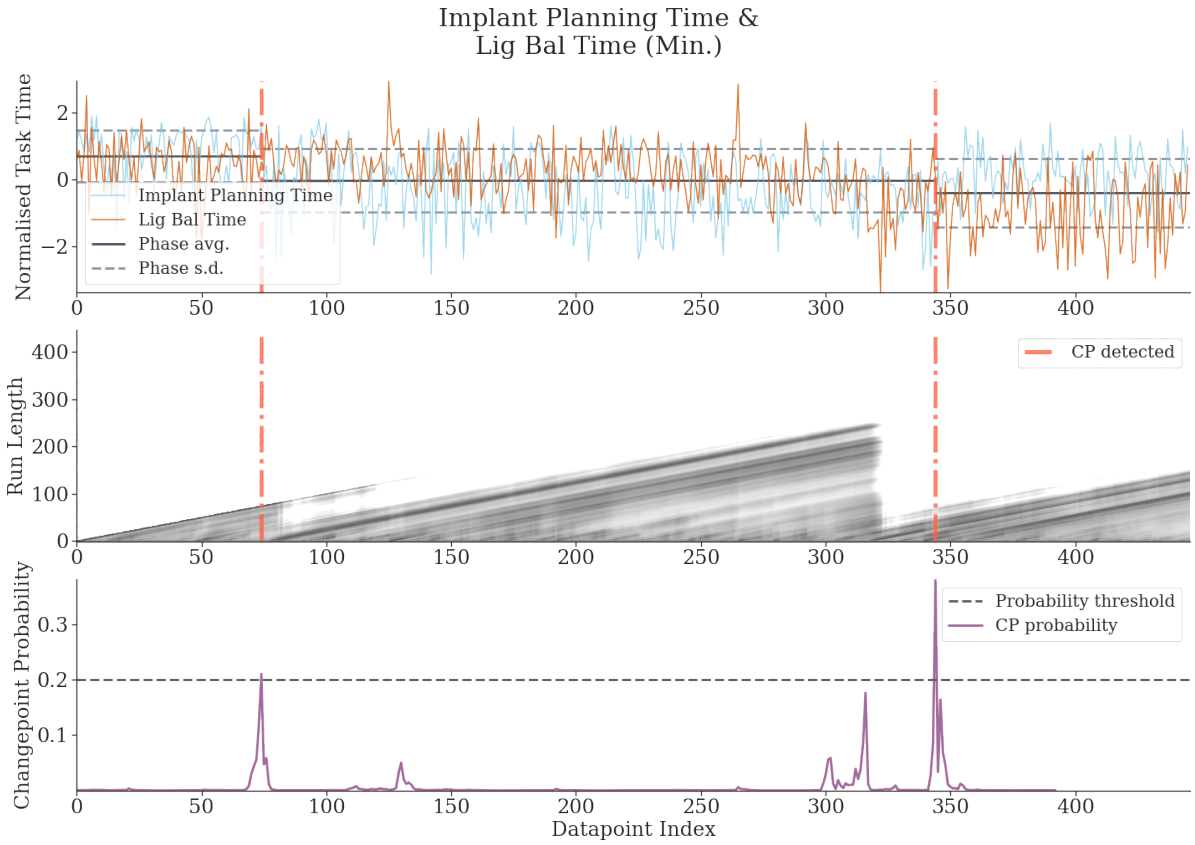


Figure 6.2: Multivariate BOCD for implant planning with ligament balancing time in minutes of Surgeon 1 post-Covid- 19. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

BOCD algorithm therefore computes the predictive probability over the current run length using a covariance matrix between of these two surgical steps. This enables the algorithm to detect a change point because the predictive probability is of course a factor inside the growth change point probabilities. It is therefore extremely beneficial to model the predictive probabilities with a covariance matrix between each data stream once a new multivariate datum is observed.

A third multivariate example selected to be tested was the ligament balancing with bone registration time due to the cumbersome nature of learning these two tasks as expressed by surgeons at the clinic. The nature of these two steps being inherently linked in difficulty adds to the curiosity of whether a golden standard truly exists. For instance, it is possible to conceive an improvement in the bone registration time. However, if this is coupled with an increase in the length of time to achieve ligament balancing then it becomes more difficult to concretely state that the surgeon had improved. As already reasoned, performing independent tasks that possess mutually inclusive outcomes would affect the surgeons learning curve.

The algorithm detects a change point between phases at 120 and 316 in Figure 6.3. Both data streams tend to move in tandem and display improvement in each surgical step over time, except for the small unexplained increase in bone registration time towards the final stages of the time series. The algorithm isolates the initial novice phase, which lasts between $1 - 119$ and is clearly distinguishable due to the many data points exceeding its standard deviation from the mean. It also assigns this change point a very high probability at near 0.6 likelihood for the beginning of a new phase. This period is also marred by a higher average phase time than the intermediate phase that succeeds it. The surgeon shows clear improvement during the phase $120 - 315$ from the initial phase. Entering into the final proficiency or expert phase the average phase time is again lower and the algorithm also assigns a high probability for a change point at 316.

Recall from Figure 5.4 the BOCD algorithm was very sensitive to the abnormally large surgical instances of bone registration time. This resulted in changepoints being detected at the latter stages of the time series,

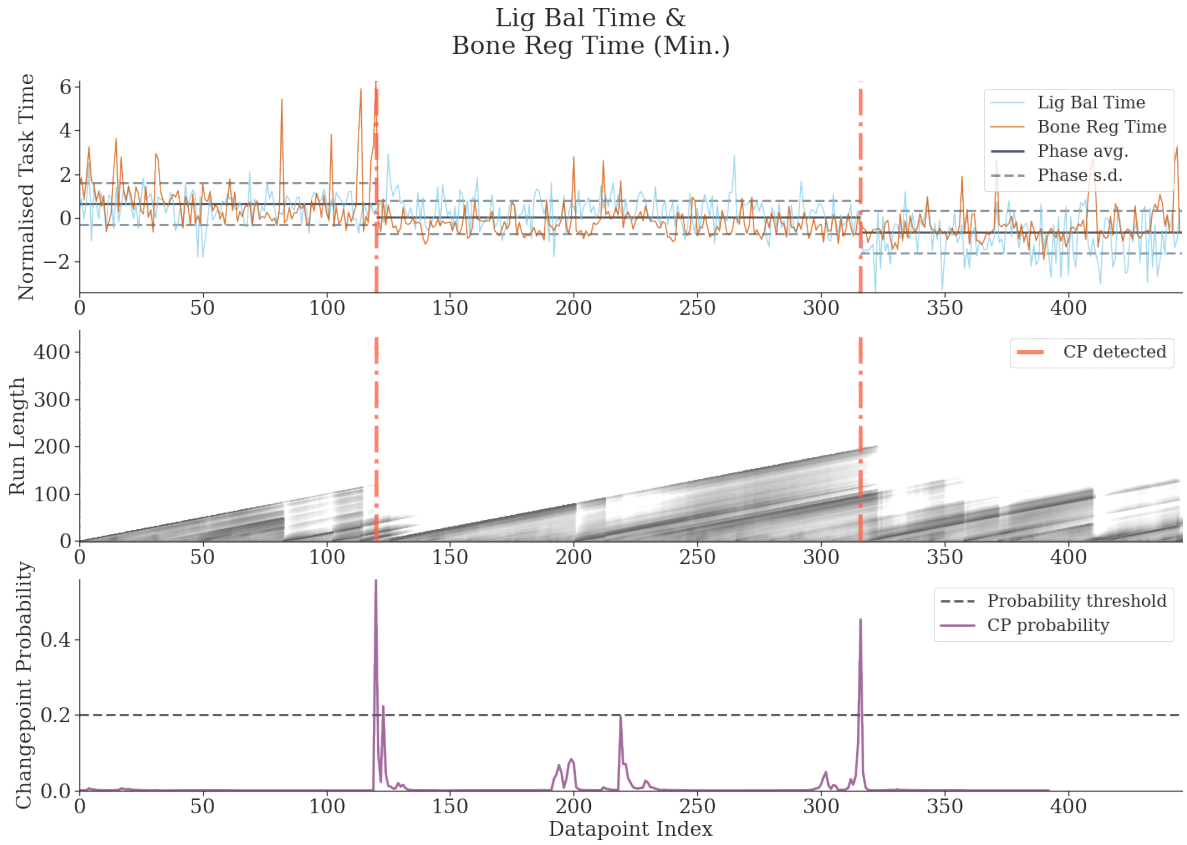


Figure 6.3: Multivariate BOCD for ligament balancing with bone registration time in minutes of Surgeon 1 post-Covid- 19. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

albeit with a small change point probability. Instead, the multivariate case is less sensitive to these data points. This forces the middle plot in Figure 6.3 to visually depict the run length ending after about 400 surgeries on a logarithmic colour scale, but in the bottom plot we observe the change point probability to be near naught. Hence despite a change occurring, BOCD had not gathered enough evidence to mark this a change point which is beneficial in our use case as we do not want to be influenced by a single surgery but rather a series of surgeries.

The final multivariate experiment involved the analysing of bone cutting with bone sawing time due to the intrinsic relationship between the two tasks. With the latter forming part of the former surgical step and thus having the ability of relating not only an improvement in time until execution of the task but also improved efficiency of the burr on versus off time. For instance, the bone sawing time may be on the decline but this could still mean little if the total cutting time is on the rise due to ineffective use of the burr.

We had seen from Figure 1.2 the burr time efficiency being on the rise year on year, signalling it is more commonplace for the surgeon to operate efficiently with the saw blade being on. Using CUSUM analysis, early inflection points were identified for each individual surgical step in Figures 4.5a, B.2a. BOCD in the univariate case identified four distinct phases in Figure 5.5 for bone sawing time, but little difference was found for bone cutting time in Figure B.4.

The algorithm detects a change point between phases at 176, 233 and 372 in Figure 6.4. Note these are identical changepoints to those identified analysing all five surgical steps in Figure 6.1. This implies the two data streams are highly influential on the learning phases of all surgical steps. At all three locations, the change point probability surpasses 0.8 meaning it is highly likely the data observed belongs to the beginning of a new phase.

The novice phase continues for $1 - 175$, exhibiting both the largest average phase time with standard deviation from the mean. Through the intermediate phase during $176 - 232$ the surgeon improves across

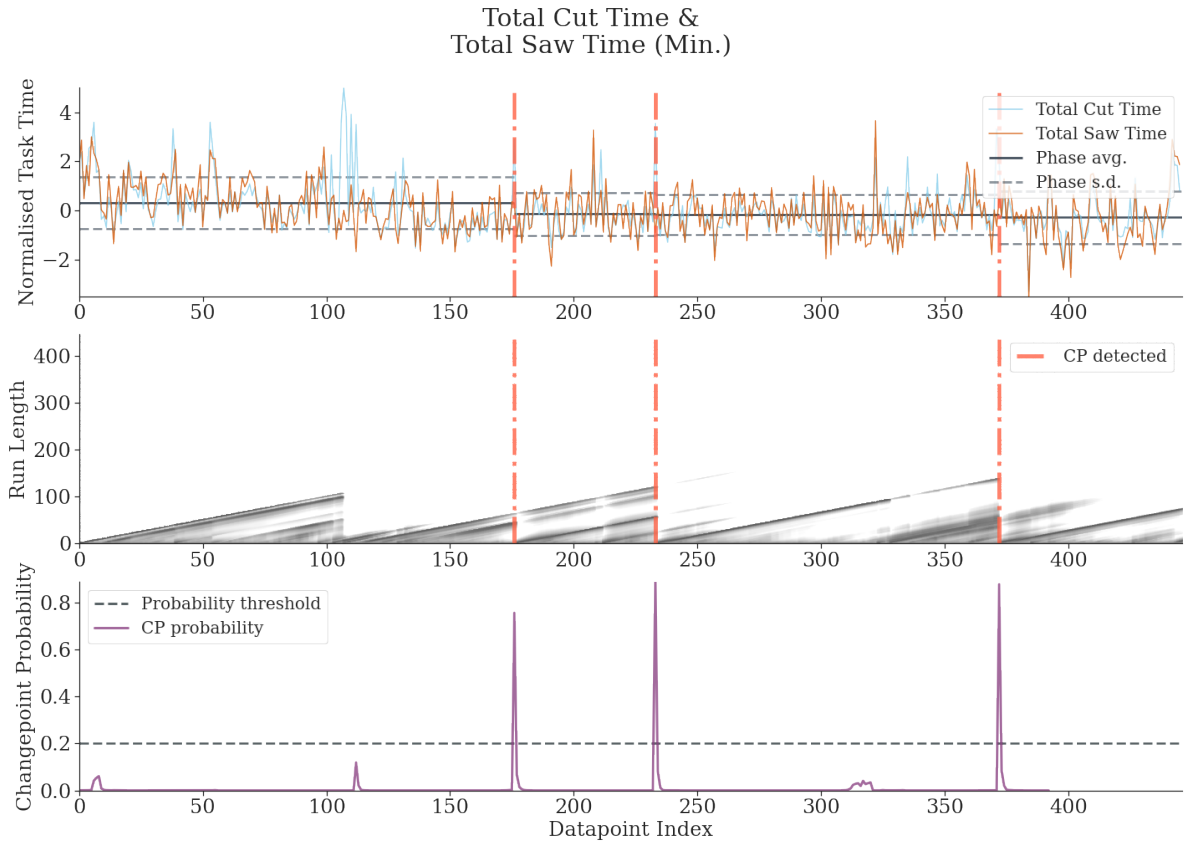


Figure 6.4: Multivariate BOCD for total cut with total saw time in minutes of Surgeon 1 post-Covid- 19. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1$, $\beta_0 = 1$, $\kappa_0 = 1$, $\mu_0 = 0$.

both tasks with both lower average time and lower standard deviation. The proficiency phase begins on 233 and arrives later than 208 with bone sawing time by itself in Figure 5.5. On the other hand, the algorithm is able to limit the redundant change point locations in Figure B.4 whilst becoming less sensitive to instances of abnormally large bone cutting time. It therefore better suited to show the surgeon gradually improving, as oppose to having an off day in surgery for whichever reason.

The final expert phase which begins at 372 has the backing of the middle plot in Figure 6.4 as to why a change point is assigned at this place. We observe that once the algorithm detects the proficient phase which begins at 233, with every newly seen data point the probability of a new run length starting after this point is very low, resulting in most of the probability mass being concentrated on the diagonal shown in darker pixels. The same effect can be seen inside the expert phase where the growth probabilities of every possible run length below the diagonal being very small. Thus indicating the run length that started at 372 most likely continues. This is an improvement on the rather indistinguishable to the naked eye growth probabilities and change point probabilities seen in Figures 5.5, B.4.

Multivariate BOCD is therefore better suited in understanding how surgical steps impact on each other. During periods when an arthoplasty surgeon exhibits both improvement and deterioration across various tasks, it can shed further detail as to how the surgeon is learning by leveraging information via the covariance matrix. This is beneficial for setting a true gold standard and is best used with specific interconnected surgical tasks, as oppose to packaging the entire surgical procedure inside a black box and simply analysing the total surgery time. An example we had shown is with implant planning and ligament balancing where the growth probabilities reset back to naught only once the algorithm had gathered enough evidence of a change occurring across both data streams.

We had also observed that uniting ligament balancing with bone registration in Figure 6.3 and bone sawing with cutting in Figure 6.4 improves the performance of the BOCD algorithm. This was achieved twofold:

by preventing anomalous data points on having a large influence on the change point location, as well as assigning clearer run length probabilities inside each middle plot.

The Bayesian approach can therefore certainly be used in answering Question 3 because *it provides another perspective to assess RAS surgery with*. By unpacking the black box-like total surgical time often used as a go to measure of improvement, the method shown here explores RAS further than Kayani et al. [30], Vermue et al. [32] and Tay et al. [31] by analysing individual surgical steps as belonging to a wider process. Where the gold standard lies remains open question and will require further comparisons with trainee surgeons to see whether they exhibit different behaviour.

Chapter 7

Experimental results of anomalous data

In this chapter we build further on the work from Section 3.7 and run experiments to see which of the offline and online methods is superior under distinct circumstances. In Sholihat et al. [37] the authors had shown that using larger initial hyperparameter values for the posterior sampled from a Student's t-distribution detects more changepoints. This is useful in the context of an early warning system because an anomalous outlier can signal a volcanic eruption is imminent. However, in the context of detecting learning curves this becomes of little use because anomalous surgeries can incorrectly signal the start of a new learning phase.

In answering Question 4 the aim is to therefore assess which method is less prone to outliers in the data, as oppose to commenting on which of the two methods can detect the most outliers. We do this by comparing the robustness of each method when there are anomalous data points, akin to a complex case with longer surgery time cropping up in amongst a time series. We show that tweaking the initial hyperparameter settings for the BOCD algorithm can match the precision accuracy of Offline BCD. However, the performance of the latter begins to falter in the multivariate case with outliers.

7.1 Univariate time series with anomalies

We had previously observed in Chapter 5 how the BOCD algorithm can perform sensitively to outlier points. This is caused due to the run length probability dropping to zero after an extreme value for a surgery time. As a result of this data point appearing, a low probability is assigned to the next data point of belonging to the previous run length since it exhibits different characteristics to the anomalous data point. This in turn forces a spike in the change point probability immediately following this anomalous data point.

It is therefore unclear whether the BOCD algorithm assigns a change point due to an accumulation in evidence from across the entire run length, or whether an anomalous point influences the spike in change point probability. Having no manner with which check this on real world data since the location of changepoints is unknown, we instead opt to test this assumption on synthetic data. The intent is to better guide the practitioners in using these algorithms in a real life surgical setting.

Following on from Section 3.7, we take the same time series data of length $N = 1117$ with limited and moderate levels of variance in the data. We populate these two sequences with 30 anomalous data point at random locations. Data with substantial level of variance is not included in our analysis because this does not match the characteristics of the surgical data we worked with. In this setting we describe anomalous as possessing a time series value double that than its immediate neighbours.

Table 7.1 shows the results of using the Offline BCD and BOCD on data with limited variance. The offline method matches the precision accuracy shown in Section 3.7 of correctly identifying the change point locations, only now this method also incorrectly identifies an additional 11 changepoints. Similarly, BOCD performs equally as well with small initial hyperparameter values and is able to identify correctly all 15 changepoints within a radius of $\delta = 10$, whilst also misidentifying an additional 7 changepoints. Both

these methods are therefore hypersensitive to anomalies when there is limited variance across segments. A possible medium is to use the BOCD method but initial hyperparameter values $\alpha_0 = \beta_0 = \kappa_0 = 1$ as these achieve a high level of precision accuracy and identify correctly 12 changepoints, whilst misidentifying zero incorrect changepoints.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	15	11	8	14
CP identified with $\delta = 10$	15	12	8	15
Incorrectly identified CP	11	0	0	7

Table 7.1: Precision accuracy comparison between Offline BCD versus BOCD with various hyperparameters. The variance of each data segment was limited and included 30 anomalous data points across all 16 segments.

In the event the variance inside the data segments is moderate we observe a small decrease in the precision accuracy in Table 7.2. This is again consistent with the results presented earlier in Section 3.7 when there were no outlier points. The difference now is that the BOCD with small initial hyperparameter values outperforms the Offline BCD method when anomalous points are introduced both in terms of precision accuracy and misidentifying less changepoints incorrectly.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	12	11	7	12
CP identified with $\delta = 10$	13	12	8	14
Incorrectly identified CP	4	1	0	3

Table 7.2: Precision accuracy comparison between Offline BCD versus BOCD with various hyperparameters. The variance of each data segment was moderate and included 30 anomalous data points across all 16 segments.

7.2 Multivariate time series with anomalies

Similarly, we wanted to test the model selection from Chapter 6 when the surgeons are assessed across several tasks simultaneously. To do this we adopt the same data sequences from Section 3.7 with limited and moderate levels of variance in the data. We proceed to randomly assign 100 anomalous data points across the three sequences that makeup the multivariate time series data.

In Table 7.3 we observe a near total collapse of the Multivariate Offline BCD method in identifying change point locations, despite a covariance structure between the time series existing. In terms of precision accuracy this algorithm identifies 5 correctly and mislabels a further 2. The Multivariate BOCD on the other hand is able to significantly outperform the offline method and correctly identifies 14 changepoints with small initial hyperparameter values. However, this comes at a high cost with also 14 incorrectly labeled changepoints. Instead a reasonable medium would be to use slightly larger initial hyperparameter values with which 12 changepoints are identified correctly and only 3 incorrectly.

Taking the multivariate data sequences with moderate variance inside the phase segments is again not suited for with the Multivariate Offline BCD when there are anomalous points. The algorithm mislabels

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	5	11	7	13
CP identified with $\delta = 10$	5	12	7	14
Incorrectly identified CP	2	3	0	14

Table 7.3: Precision accuracy comparison between Multivariate Offline BCD versus Multivariate BOCD with various hyperparameters. The variance of each data segment was limited and included 100 anomalous data points across all 16 segments.

incorrect locations as changepoints more often than correctly identifying the real changepoints. The Multivariate BOCD is able to achieve a high level of precision accuracy under two different initial hyperparameter settings. It outperforms the offline method across all three cases we had tested.

Precision	Method			
	Offline	Online $\alpha_0 = 1$ $\beta_0 = 1$ $\kappa_0 = 1$	Online $\alpha_0 = 5$ $\beta_0 = 5$ $\kappa_0 = 5$	Online $\alpha_0 = 0.1$ $\beta_0 = 0.1$ $\kappa_0 = 0.1$
CP identified with $\delta = 5$	2	14	10	12
CP identified with $\delta = 10$	3	14	10	14
Incorrectly identified CP	4	2	0	7

Table 7.4: Precision accuracy comparison between Multivariate Offline BCD versus Multivariate BOCD with various hyperparameters. The variance of each data segment was moderate and included 100 anomalous data points across all 16 segments.

7.3 Summary

In earlier chapters throughout this thesis a question which cropped up was to what extent do outliers impact on the findings of learning phases. Working on synthetic data, we have now provided a set of results that help form a best practice approach of working with Bayesian methods for segmentation of time series with anomalous data points.

The results show that with data from a univariate time series sequence the Offline BCD method continues to outperform the BOCD algorithm. Comparable results can be achieved for the latter method by choosing smaller initial hyperparameter values. This follows the theoretical observations from Section 3.5 and Section 3.7, where it was pointed out that smaller hyperparameter values result in less dispersed posterior distribution of run lengths and thus more changepoints being detected.

On the other hand, choosing large hyperparameter values is an antithetical example to the MAP estimate derivations from Section 2.2. If the hyperparameters are poorly chosen, the resulting prior distribution may not accurately capture our beliefs about the true parameters, and this can lead to incorrect or biased posterior estimates. We observe that the BOCD algorithm achieves the lowest precision accuracy for correctly identifying changepoints with larger hyperparameter values.

With multivariate time series analysis, the Offline BCD method should not be used as it poorly identifies changepoints in data sequences with outliers. The online BOCD method achieves a near perfect precision accuracy by detecting 14 changepoints correctly. This method can however perform sensitively to outliers

and therefore initial hyperparameter values are an important choice.

We therefore answer Question 4 by stating that *offline methods are more suited for univariate data but online methods possess superior performance in the multivariate case.* Furthermore, in order to avoid miscommunication for the beginning of a new learning phases of surgeons, a good medium is to use initial hyperparameter settings which do not capture all changepoints correctly but that are also less sensitive to outliers.

Chapter 8

Conclusion

In this chapter we summarise the findings to the four questions posed in this thesis, as well as to provide a comparison between the different change point detection methods used. We also discuss the future possibilities in building supportive surgical tools with data science.

8.1 Research questions

In this thesis, we had set out to provide surgeons in the clinic with concrete performance metrics by means of analysing surgical data. On the way, we compared the three change point methods, CUSUM, Offline BCD and BOCD, and evaluated their performance in solving for the four clinical questions at hand.

In answering Question 1, we had shown that learning curves exist for Surgeon 1 in performing TKA surgery, as well as for the five surgical subtasks. By working with a greater amount of data in relative terms than [30, 31, 32], we had been able to go further and to find learning curves for ligament balancing and bone registration time. However, using CUSUM analysis with more data did provide challenges. We had found the learning curves to be much longer in number of surgeries than in previous literature findings. It is also not possible to simultaneoulsy compare multivariate data sequences. It was therefore concluded that this method is not useful to compare data of invariant lengths, particularly for comparing new surgeons who may be starting with more experienced surgeons.

We had discussed both the Offline BCD and BOCD as alternative change point detection methods and thus expanded on the notion of learning curves instead consisting of multiple phases. Due to comparative results being obtained on synthetic data for the offline and online Bayesian methods, the decision was made to work with BOCD. The reasoning follows that as part of building supportive surgical tools, surgeons would benefit most if feedback was provided in real time. As opposed to collecting all the data and looking back retrospectively at what happened, as is done with offline methods such as Offline BCD and CUSUM.

Working with real world Mako RAS data, it was shown for multiple phases to exist when using BOCD. Thus in answering Question 2 it is true Surgeon 1 continuously improves beyond the change points of the first learning phase, again disproving other literature findings in TKA surgery. This method is therefore better equipped to communicate to the clinic when the surgeons performance both improves and deteriorates, such as in the period during which the surgery was shut due to the Covid - 19 pandemic.

To streamline feedback being made to surgeons, surgical subtasks had also been analysed simultaneously using Multivariate BOCD, with which we were able to model correlations between the features with the use of a covariance matrix. Thus simplifying the complexity exacerbated by the multitude of tasks in surgery, as well as the black box - like total surgical time, we showed how interconnected surgical subtasks impacted on completion time of one another. In answering Question 3 we had therefore shown where the gold standard is for the most difficult tasks such as ligament balancing and bone registration time, as well as for more analog tasks such as bone sawing and cutting time.

A common complication which arose in answering both questions with Bayesian change point detection is whether this method was sensitive to jumps in the time series data. In Question 4 we had therefore assessed whether our choice of method was in fact correct for use with surgical data. We ran experiments on synthetic data and showed both the offline and online methods provide comparative results in the univariate case when the correct set of initial hyperparameters was used for the latter. The performance of both also decreased with increased variance in the data sequences. However, with the introduction of anomalous data points, the offline method overcompensates by detecting too many instances of changepoints and sees a drop in its precision as a result. The suitability of the online method in change point detection is that we do not have too many under BOCD. It was concluded for analyses of data in surgery, where complications result in jumps in the time series, it is best to use BOCD as this method is less sensitive to outliers.

8.2 Limitations and future work

The work presented here looks promising but more can be done in data science for surgery. For instance, further analysis are required to incorporate bone alignment of the knee, care quality and patient insights with data science. Data from other surgeons in the clinic will also need to be reviewed, with Bayesian priors from the gold standard of Surgeon 1 being used to precisely identify the learning phases of junior surgeons.

The issues surrounding the slow implementation of alignment data inside data scientific tools is associated with the resource constraints surrounding much needed input from surgical professionals. Overrides to bone alignment rules can take place from one patient to the next. There is also a lack of consensus in the industry regarding the best knee alignment, with different alignment strategies existing across nations, something that would prove difficult if had to be hard coded inside an algorithm.

Any future supportive surgical tools will therefore necessitate a link to be made between the expertise of a human surgeon to the decision on knee alignment being made. What form this intelligence inside surgical tools will take remains to be seen. Questions will also have to be asked whether the recommended settings of using the Mako RAS will match future cohorts of patients.

The added effect of simply looking at care quality can be misleading for the reason of patients answering questions to best suit their situation, as oppose to what surgical professionals believe best suits the recovery of a patient. Extra care is therefore required when communicating conclusions from any data related to patient behaviour as those who are in discomfort most tend to disapprove most often too.

In this thesis we had only worked time series data. In reality, there is a need to tap into all available data sources in order to provide a diverse, effective and beneficial supportive surgical tools with data science. To achieve the aforementioned goals will require time and effort from both ends of the spectrum - surgeons and data scientists.

Bibliography

[1] J.H. Lasater. *Yogabody: Anatomy, Kinesiology, and Asana*. Rodmell Press, 2009. ISBN: 9781930485235.

[2] W. J. Hozack. "Multicenter Analysis of Outcomes after Robotic-Arm Assisted Total Knee Arthroplasty". In: *Orthopaedic Proceedings* 100-B.SUPP_12 (2018), pp. 38–38. DOI: 10.1302/1358-992X.2018.12.038.

[3] Y.N. Harari. *Sapiens: A Brief History of Humankind*. HarperCollins, 2015. ISBN: 9780062316103.

[4] Emily L. Hampp et al. "Robotic-Arm Assisted Total Knee Arthroplasty Demonstrated Greater Accuracy and Precision to Plan Compared with Manual Techniques". In: *J. Knee Surg.* 32.3 (2019), pp. 239–250. DOI: 10.1055/s-0038-1641729.

[5] Robert C. Marchand et al. "Patient Satisfaction Outcomes after Robotic Arm-Assisted Total Knee Arthroplasty: A Short-Term Evaluation". In: *J. Knee Surg.* 30.09 (2017), pp. 849–853. DOI: 10.1055/s-0037-1607450.

[6] B. Kayani et al. "Robotic-arm assisted total knee arthroplasty". In: *Bone & Joint Journal* 100-B.7 (2018), p. 930. DOI: 10.1302/0301-620X.100B7.BJJ-2017-1449.R1.

[7] Cécile Batailler et al. "MAKO CT-based robotic arm-assisted system is a reliable procedure for total knee arthroplasty: a systematic review". In: *Knee Surg. Sports Traumatol. Arthrosc.* 29.11 (2021), pp. 3585–3598. DOI: 10.1007/s00167-020-06283-z.

[8] Steven Kurtz et al. "Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030". In: *J. Bone Joint Surg. Am.* 89.4 (2007), pp. 780–785. DOI: 10.2106/JBJS.F.00222.

[9] D. Culliford et al. "Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink". In: *Osteoarthritis Cartilage* 23.4 (2015), pp. 594–600. DOI: 10.1016/j.joca.2014.12.022.

[10] Alexander Klug et al. "The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years". In: *Knee Surg. Sports Traumatol. Arthrosc.* 29.10 (2021), pp. 3287–3298. DOI: 10.1007/s00167-020-06154-7.

[11] Dursun Delen and Sudha Ram. "Research challenges and opportunities in business analytics". In: *Journal of Business Analytics* (2018). DOI: 10.1080/2573234X.2018.1507324.

[12] Lena Maier-Hein et al. "Surgical data science for next-generation interventions". In: *Nat. Biomed. Eng.* 1 (2017), pp. 691–696. DOI: 10.1038/s41551-017-0132-7.

[13] D. J. Power et al. "Defining business analytics: an empirical approach". In: *Journal of Business Analytics* (2018). DOI: 10.1080/2573234X.2018.1507605.

[14] *Inside Team Sky*. [Online; accessed 21. Oct. 2022]. 2016. URL: https://www.cyclist.co.uk/team-ineos-grenadiers/1524/inside-team-sky.

[15] *How Team Sky took on the world*. [Online; accessed 21. Oct. 2022]. 2018. URL: https://www.bbc.com/sport/cycling/21331484.

[16] William Fotheringham. "Military precision forms the tip of Team Sky's cycling iceberg". In: *the Guardian* (2010). URL: https://www.theguardian.com/sport/2010/jul/18/team-sky-tour-de-france.

[17] Syed F. H. Shah and Zach Sheridan. "When predictive analytics goes wrong: what can healthcare learn from Formula 1?" In: *Br. J. Hosp. Med.* (2020). DOI: 10.12968/hmed.2020.0389.

[18] Nicole Powell-Dunford et al. "Transferring Aviation Practices into Clinical Medicine for the Promotion of High Reliability". In: *Aerosp. Med. Hum. Perform.* 88.5 (2017), pp. 487–491. DOI: 10.3357/AMHP.4736.2017.

[19] Pierre Wannaz. *Big data for pilots: how to exploit all the potential of this source of information to make it a real game changer in pilots' awareness and performance improvement - CEFA Aviation*. [Online; accessed 18. Oct. 2022]. 2021. URL: https://www.cefa-aviation.com/big-data-for-pilots-a-nice-to-have-or-a-real-game-changer.

[20] Wei Tian et al. "Flight maneuver intelligent recognition based on deep variational autoencoder network". In: *EURASIP J. Adv. Signal Process.* 2022.1 (2022), pp. 1–23. DOI: 10.1186/s13634-022-00850-x.

[21] Aidan Mcparland, Alun Ackery, and Allan S. Detsky. "Advanced analytics to improve performance: can healthcare replicate the success of professional sports?" In: *BMJ Qual. Saf.* 29.5 (2020), pp. 405–408. DOI: 10.1136/bmjqs-2019-010415.

[22] *Orthopedic Surgical Robots Market Size, Industry Outlook, 2030*. [Online; accessed 10. Oct. 2022]. 2022. URL: https://www.strategicmarketresearch.com/market-report/orthopedic-surgical-robots-market.

[23] Asuka Takai et al. "Bayesian Estimation of Potential Performance Improvement Elicited by Robot-Guided Training". In: *Front. Neurosci.* 0 (2021). DOI: 10.3389/fnins.2021.704402.

[24] Sutuke Yibulayimu et al. "An explainable machine learning method for assessing surgical skill in liposuction surgery". In: *Int. J. CARS* (2022), pp. 1–12. DOI: 10.1007/s11548-022-02739-4.

[25] Seung-Kook Jun et al. "Evaluation of robotic minimally invasive surgical skills using motion studies". In: *J. Robot. Surg.* 7.3 (2013), pp. 241–249. DOI: 10.1007/s11701-013-0419-y.

[26] Hassan Ismail Fawaz et al. "Automatic Alignment of Surgical Videos Using Kinematic Data". In: Springer International Publishing, 2019, pp. 104–113. DOI: 10.1007/978-3-030-21642-9_14.

[27] Ziheng Wang and Ann Majewicz Fey. "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery". In: *Int. J. CARS* 13.12 (2018), pp. 1959–1970. DOI: 10.1007/s11548-018-1860-1.

[28] Hassan Ismail Fawaz et al. "Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks". In: *Int. J. CARS* 14.9 (2019), pp. 1611–1617. DOI: 10.1007/s11548-019-02039-4.

[29] Jingyu Jiang et al. "Evaluation of robotic surgery skills using dynamic time warping". In: *Comput. Methods Programs Biomed.* 152 (2017), pp. 71–83. DOI: 10.1016/j.cmpb.2017.09.007.

[30] Babar Kayani et al. "Robotic-arm assisted total knee arthroplasty has a learning curve of seven cases for integration into the surgical workflow but no learning curve effect for accuracy of implant positioning". In: *Knee Surg. Sports Traumatol. Arthrosc.* 27.4 (2019), p. 1132. DOI: 10.1007/s00167-018-5138-5.

[31] Mei Lin Tay et al. "Robotic-arm assisted total knee arthroplasty has a learning curve of 16 cases and increased operative time of 12 min". In: *ANZ J. Surg.* 92.11 (2022), pp. 2974–2979. DOI: 10.1111/ans.17975.

[32] Hannes Vermue et al. "Robot-assisted total knee arthroplasty is associated with a learning curve for surgical time but not for component alignment, limb alignment and gap balancing". In: *Knee Surg. Sports Traumatol. Arthrosc.* 30.2 (2022), pp. 593–602. DOI: 10.1007/s00167-020-06341-6.

[33] Paul Fearnhead. "Exact Bayesian curve fitting and signal segmentation". In: *IEEE Trans. Signal Process.* 53.6 (2005), pp. 2160–2166. DOI: 10.1109/TSP.2005.847844.

[34] Paul Fearnhead. "Exact and efficient Bayesian inference for multiple changepoint problems". In: *Statist. Comput.* 16.2 (2006), pp. 203–213. DOI: 10.1007/s11222-006-8450-8.

[35] Ryan Prescott Adams and David J. C. MacKay. "Bayesian Online Changepoint Detection". In: *arXiv* (2007). DOI: 10.48550/arXiv.0710.3742.

[36] Zhaohui Wang et al. *Online Changepoint Detection on a Budget*. IEEE Computer Society, 2021. DOI: 10.1109/ICDMW53433.2021.00057.

[37] Seli Siti Sholihat, Sapto Wahyu Indratno, and Utriweni Mukhaiyar. "The role of parameters in Bayesian Online Changepoint Detection: detecting early warning of mount Merapi eruptions". In: *Heliyon* 7.7 (2021). DOI: 10.1016/j.heliyon.2021.e07482.

[38] *Mako TKA Surgical Guide*. [Online; accessed 12. Mar. 2023]. 2016. URL: https://www.strykermeded.com/medical-devices/robotics-navigation/robotics-navigation/mako-tka/.

[39] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer. ISBN: 9780387310732.

[40] Norman Fenton, Martin Neil, and Daniel Berger. "Bayes and the Law". In: *Annu. Rev. Stat. Appl.* 3 (2016), pp. 51–77. DOI: 10.1146/annurev-statistics-041715-033428.

[41] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009. ISBN: 9780387848587.

[42] Ronald Fisher. *On the mathematical foundations of theoretical statistics*. 1922. DOI: 10.1098/rsta.1922.0009.

[43] Edwin J. G. Pitman. "Sufficient statistics and intrinsic accuracy". In: *Math. Proc. Cambridge Philos. Soc.* 32.4 (1936), pp. 567–579. DOI: 10.1017/S0305004100019307.

[44] *Can 10,000 hours of practice make you an expert?* [Online; accessed 14. Dec. 2023]. 2014. URL: https://www.bbc.co.uk/news/magazine-26384712.

[45] Leonard E. Baum et al. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". In: *Ann. Math. Stat.* 41.1 (1970), pp. 164–171. DOI: 10.1214/aoms/1177697196.

[46] Kevin Murphy. "Conjugate Bayesian analysis of the Gaussian distribution". In: (2007).

[47] Xiang Xuan and Kevin Murphy. "Modeling changing dependency structure in multivariate time series". In: (2007), pp. 1055–1062. DOI: 10.1145/1273496.1273629.

[48] Clemens Schopper et al. "The learning curve in robotic assisted knee arthroplasty is flattened by the presence of a surgeon experienced with robotic assisted surgery". In: *Knee Surg. Sports Traumatol. Arthrosc.* 31.3 (2022), pp. 760–767. DOI: 10.1007/s00167-022-07048-6.

Appendix A

Bayesian estimation

Conjugate priors are useful because they allow for computational efficiency when incorporating the prior inside the posterior. If the prior and the posterior belong to the same distribution, then they are conjugate to each other. The general problem can then be solved in a closed form expression form. We consult the work outlined in [46] to answer why the natural conjugate prior is of the form :

$$\mathbb{P}(\mu) = \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \tag{A.1}$$

Recall from Equation 2.11 that the likelihood function is of the form:

$$\begin{aligned}\mathbb{P}(\mathbf{x}|\mu,\sigma) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right)\end{aligned} \tag{A.2}$$

Utilising the MLE technique from Equation 2.9 to define the empirical mean and variance over the sample observations:

$$\hat{\mu}_{\mathrm{ML}} = \frac{1}{n}\sum_i x_i \tag{A.3}$$

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \tag{A.4}$$

These are simply the sample mean $\hat{\mu}_{\mathrm{ML}}$ and sample variance $\hat{\sigma}^2_{\mathrm{ML}}$ we already derived in Equations 2.13-2.14. Then the quadratic term inside the exponent can instead be written as:

$$\begin{aligned}\sum_i (x_i-\mu)^2 &= \sum_i \left((x_i - \hat{\mu}_{\mathrm{ML}}) - (\mu - \hat{\mu}_{\mathrm{ML}})\right)^2 \\ &= \sum_i (x_i - \hat{\mu}_{\mathrm{ML}})^2 - 2\sum_i (x_i - \hat{\mu}_{\mathrm{ML}})(\mu - \hat{\mu}_{\mathrm{ML}}) + \sum_i (\hat{\mu}_{\mathrm{ML}} - \mu)^2 \\ &= n\hat{\sigma}^2_{\mathrm{ML}} + n(\hat{\mu}_{\mathrm{ML}} - \mu)^2\end{aligned} \tag{A.5}$$

Since:

$$2\sum_i (x_i - \hat{\mu}_{\mathrm{ML}})(\mu - \hat{\mu}_{\mathrm{ML}}) = 2(n\hat{\mu}_{\mathrm{ML}} - n\hat{\mu}_{\mathrm{ML}})(\mu - \hat{\mu}_{\mathrm{ML}}) = 0 \tag{A.6}$$

We can the rewrite Equation A.2 for the likelihood as being:

$$\begin{aligned}\mathbb{P}\left(\mathbf{x}|\mu,\sigma\right) &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(-\frac{1}{2\sigma^2}\left(n\hat{\sigma}^2_{\mathrm{ML}}+n\left(\hat{\mu}_{\mathrm{ML}}-\mu\right)^2\right)\right) \\ &\propto \frac{1}{\sigma^n}\exp\left(-\frac{n\hat{\sigma}^2_{\mathrm{ML}}}{2\sigma^2}\right)\exp\left(-\frac{n}{2\sigma^2}\left(\hat{\mu}_{\mathrm{ML}}-\mu\right)^2\right)\end{aligned} \tag{A.7}$$

Then if σ^2 is a constant, we only keep the exponent with term μ for the likelihood:

$$\begin{aligned}\mathbb{P}\left(\mathbf{x}|\mu\right) &= \exp\left(-\frac{n}{2\sigma^2}\left(\hat{\mu}_{\mathrm{ML}}-\mu\right)^2\right) \\ &\propto \mathcal{N}\left(\hat{\mu}_{\mathrm{ML}}|\mu,\frac{\sigma^2}{n}\right)\end{aligned} \tag{A.8}$$

Therefore to simplify the derivation for the posterior, we take the prior distribution to be conjugate to the likelihood function:

$$\begin{aligned}\mathbb{P}(\mu) &= \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) \\ &\propto \mathcal{N}\left(\mu|\mu_0,\sigma_0^2\right)\end{aligned} \tag{A.9}$$

Which is simply Equation 2.16 with the constant term dropped. The posterior distribution is then:

$$\mathbb{P}(\mu|\mathbf{x}) = \mathcal{N}\left(\mu|\mu_{\mathrm{MAP}},\sigma^2_{\mathrm{MAP}}\right) \tag{A.10}$$

Through multiplication of the likelihood with the prior as we had done in Equation 2.17 and simple manipulation we show that it is indeed the case that the posterior takes the form of another Gaussian process with parameters $\left[\mu_{\mathrm{MAP}},\sigma^2_{\mathrm{MAP}}\right]$:

$$\begin{aligned}\mathbb{P}(\mu|\mathbf{x}) &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(-\frac{1}{2\sigma^2}\sum_i\left(x_i-\mu\right)^2\right)\cdot\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) \\ &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n\cdot\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{1}{2\sigma^2}\sum_i\left(x_i-\mu\right)^2-\frac{1}{2\sigma_0^2}\left(\mu-\mu_0\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_i\left(x_i-\mu\right)^2-\frac{1}{2\sigma_0^2}\left(\mu-\mu_0\right)^2\right) \\ &= \exp\left(-\frac{1}{2\sigma^2}\sum_i\left(x_i^2-2x_i\mu+\mu^2\right)-\frac{1}{2\sigma_0^2}\left(\mu^2-2\mu\mu_0+\mu_0^2\right)\right)\end{aligned} \tag{A.11}$$

We then separate this expression into terms that depend and those that are independent of μ:

$$\mathbb{P}(\mu|\mathbf{x}) = \exp\left(\underbrace{-\frac{\mu^2}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}_{a\mu^2}+\underbrace{\mu\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_i x_i}{\sigma^2}\right)}_{b\mu}-\underbrace{\left(\frac{\mu_0^2}{2\sigma_0^2}+\frac{\sum_i x_i^2}{2\sigma^2}\right)}_{c}\right) \tag{A.12}$$

Bishop [39, p.86, p.98] solves for μ with *completing the square* by matching first and second order powers of μ, whilst Murphy [46] takes advantage of the fact that any quadratic polynomial can be written simply as:

$$p^2-2pq+q^2=(p-q)^2 \tag{A.13}$$

Then Equation A.12 can be rewritten:

$$\mathbb{P}(\mu|\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2_{\text{MAP}}}\left(\mu - \mu_{\text{MAP}}\right)^2\right) \tag{A.14}$$

Where we match the coefficients of a from inside Equation A.12 to solve for the posterior variance σ^2_{MAP}:

$$\begin{aligned} -\frac{\mu^2}{2\sigma^2_{\text{MAP}}} &= -\frac{\mu^2}{2}\left(\frac{1}{\sigma^2_0} + \frac{n}{\sigma^2}\right) \\ \frac{1}{\sigma^2_{\text{MAP}}} &= \frac{1}{\sigma^2_0} + \frac{n}{\sigma^2} \\ \sigma^2_{\text{MAP}} &= \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma^2_0}} \end{aligned} \tag{A.15}$$

Similarly, matching the coefficients of b from inside Equation A.12 to solve for the posterior mean μ_{MAP}:

$$\begin{aligned} -\frac{\mu\mu_{\text{MAP}}}{\sigma^2_{\text{MAP}}} &= \mu\left(\frac{\sum_i x_i}{\sigma^2} + \frac{\mu_0}{\sigma^2_0}\right) \\ \frac{\mu_{\text{MAP}}}{\sigma^2_{\text{MAP}}} &= \frac{\sum_i x_i}{\sigma^2} + \frac{\mu_0}{\sigma^2_0} \\ &= \frac{n\sigma^2_0\hat{\mu}_{\text{ML}} + \sigma^2\mu_0}{\sigma^2\sigma^2_0} \end{aligned} \tag{A.16}$$

Placing μ_n on one side:

$$\begin{aligned} \mu_{\text{MAP}} &= \frac{n\sigma^2_0}{n\sigma^2_0 + \sigma^2}\hat{\mu}_{\text{ML}} + \frac{\sigma^2}{n\sigma^2_0 + \sigma^2}\mu_0 \\ &= \sigma^2_{\text{MAP}}\left(\frac{\mu_0}{\sigma^2_0} + \frac{n\hat{\mu}_{\text{ML}}}{\sigma^2}\right) \end{aligned} \tag{A.17}$$

Finally, the posterior distribution takes the same functional form as the prior, since the latter is conjugate to the likelihood function, and with parameters $\left[\mu_{\text{MAP}}, \sigma^2_{\text{MAP}}\right]$:

$$\begin{aligned} \mathbb{P}(\mu|\mathbf{x}) &= \mathbb{P}(\mathbf{x}|\mu)\mathbb{P}(\mu) \\ &\propto \mathcal{N}(\mu|\mu_{\text{MAP}}, \sigma^2_{\text{MAP}}) \end{aligned} \tag{A.18}$$

Appendix B

Additional results

CUSUM of implant planning time

The Mako RAS system enables the surgeon to perform pre-operative implant planning using the patient specific bone model and implant templates built on top of the CT scan. The primary purpose of pre-operative implant planning is to size, align, and position the implant specifically to each patients bony anatomy. Fine tuning of the implant plan using additional clinical information such as patient specific kinematics, fixed deformities, and soft tissue tension will be completed intraoperatively.

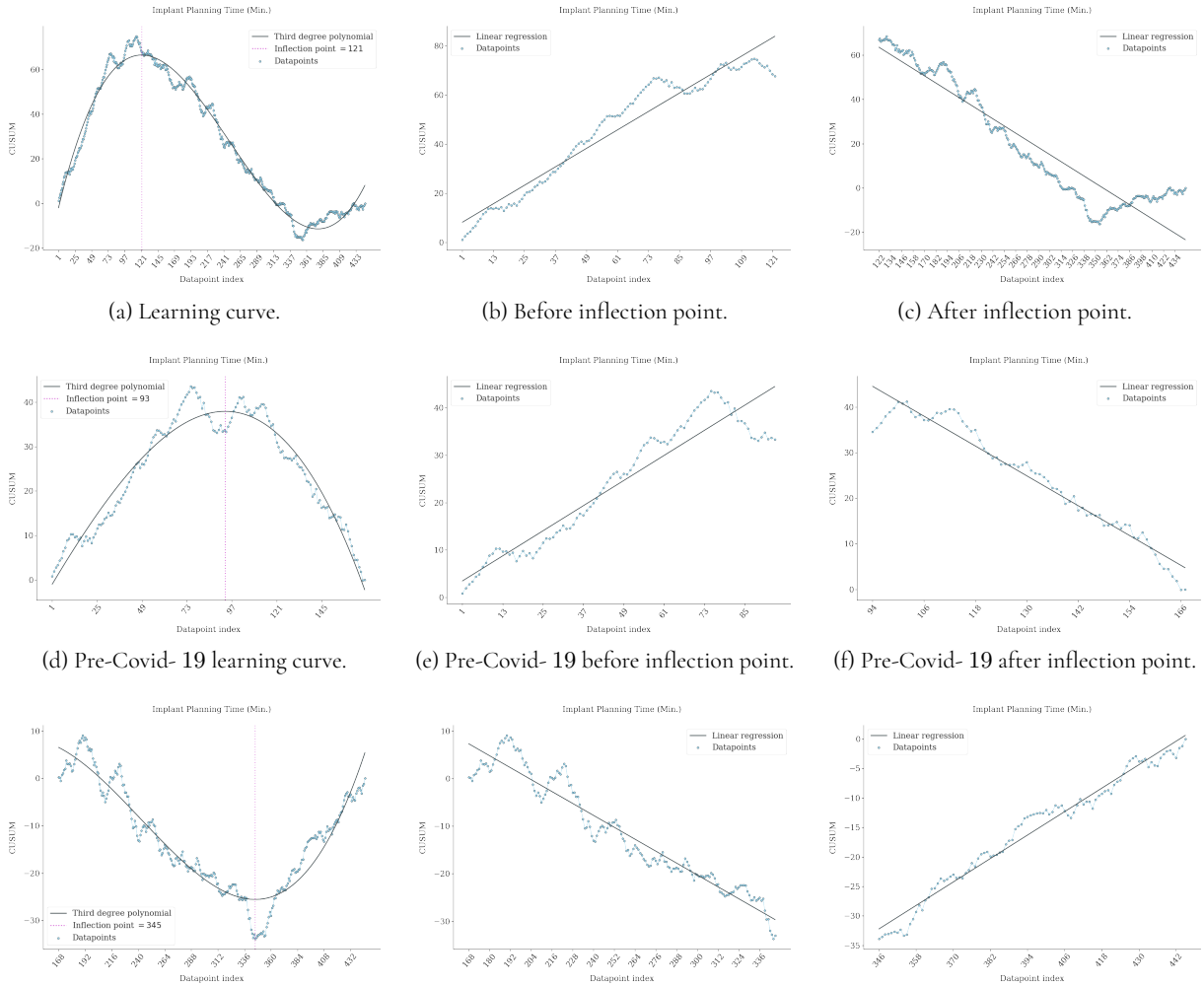
The implant planning surgical step is a balancing trick between experience in using the RAS system, knowledge of distinctive anatomical structures, the guile together with speed to adjust any pre-surgical plan intraoperatively without this impacting on the total surgical time and at the same time using the pre-surgical time as efficiently as possible by performing other tasks in the patient pathway care. A well prepared implant plan therefore does not go unnoticed.

The inflection point for the learning curve in Figure B.1a shows 121. The inexperienced phase is short and the curve is steep. Furthermore, an inflection point is found at 107 in Figure B.1d if we take only the first 167 consecutive surgeries that took place before the pandemic, but no learning curve is identified post-Covid - 19 in Figure B.1g. Once again indicating that Surgeon 1 does not require much time to become proficient in performing the pre-surgical implant planning.

CUSUM of cutting time

The bone cutting phase prepares the knee for the implant positioning by performing resections in the bone and it encompasses the active and inactive modes of the saw blade, as well as the switch between cut types. This on-off mechanism is crucially what determines the speed and proficiency with which the surgeon is able to perform the bone sawing. Many stop start motions will in turn increase the time of this surgical step.

Surgeon 1 is able to quickly hone the skills to perform the bone cutting step with a steep inflection point being found at the 143 surgery mark in Figure B.2a. The pre-Covid- 19 data gives off an inflection point at 107 in Figure B.2d, despite having a worse linear regression fit to the data in Figures B.2e , B.2f. Albeit an inflection point is found at 136 (304 consecutive surgery) in Figure B.2g, the learning curve is flattened and remains centered around zero, thus lacking the conviction of the surgeon improving during the post-Covid - 19 period.



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

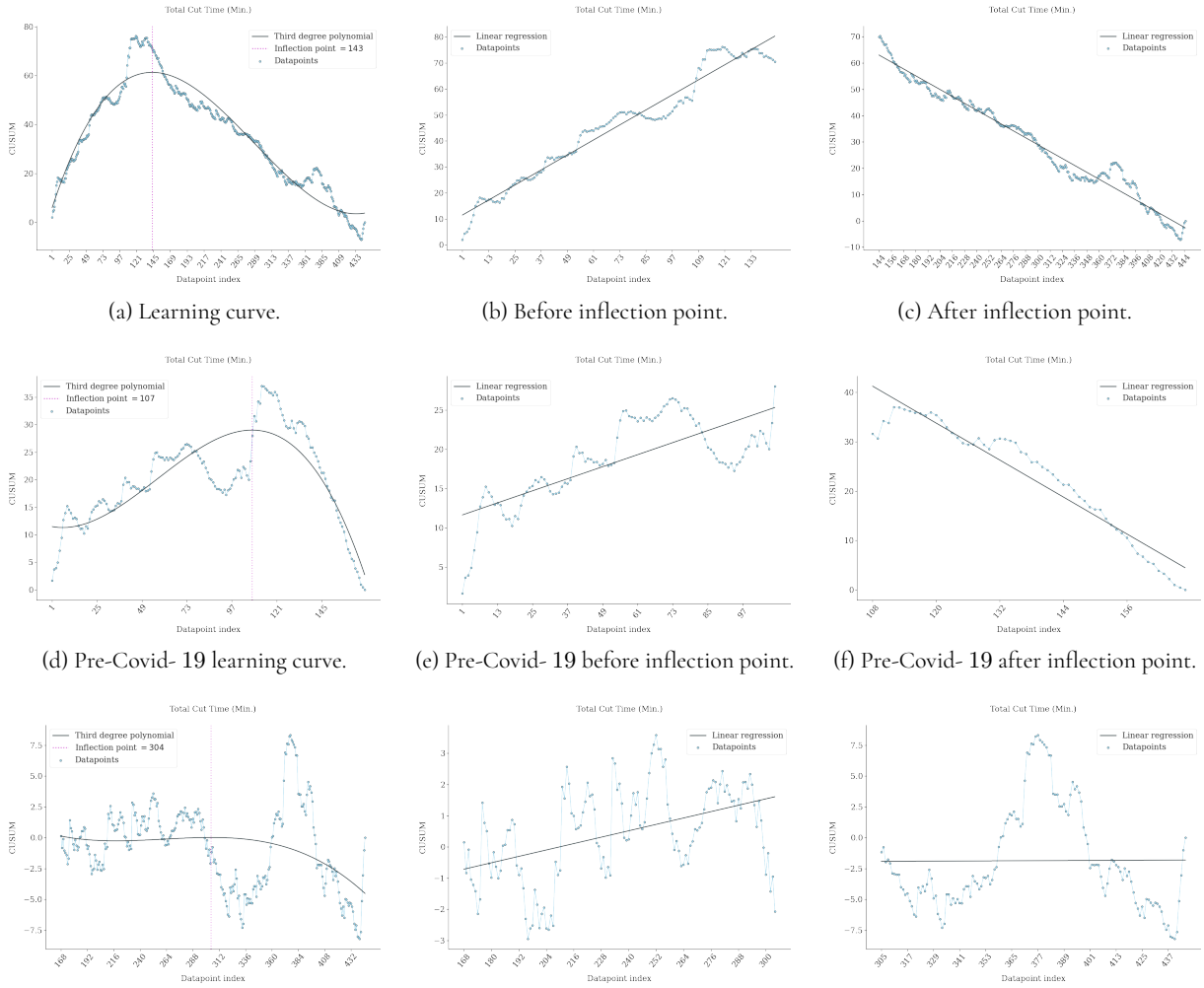
Figure B.1: CUSUM analysis for the implant planning time in minutes of Surgeon 1.

BOCD of implant planning time

For the time series of the implant planning time BOCD identifies changepoints at 100, 112, 355, 363 and 364. Due to the probabilistic nature of the algorithm, a change point can be assigned very soon after already designating one as it continues to gather more evidence from the data. For simplicity here we opt to forego changepoints which occur consecutively near each other to best fit our predefined surgical phases. Observing from the middle plot in Figure B.3 which run length possess the lowest logarithmic colour scale and thus should be merged, we assume the three defined change point locations are therefore 100, 112 and 360.

The novice together with the intermediate phase persist until the change point into the proficient phase occurs at 112. This is only a handful of surgeries sooner than the inflection point of the learning curve at 121 with CUSUM in Figure B.1a. There is a significantly evident drop in the average phase time from the start of the novice phase to when surgeon enters proficiency, as well as the surgeon being able to standardise the time to perform this surgical step with a more concentrated standard deviation around the mean.

The final stage that takes place between $360 - 446$ is the expert phase and should by definition show an improvement on the proficient phase. However, the change point at 360 takes place because the surgeons performance time increases on average. The change point probability of over 0.3 at that location is the



(a) Learning curve. (b) Before inflection point. (c) After inflection point.

(d) Pre-Covid- 19 learning curve. (e) Pre-Covid- 19 before inflection point. (f) Pre-Covid- 19 after inflection point.

(g) Post-Covid- 19 learning curve. (h) Post-Covid- 19 before inflection point. (i) Post-Covid- 19 after inflection point.

Figure B.2: CUSUM analysis for the total cutting time in minutes of Surgeon 1.

largest for this time series and signal that the data post 360 surgeries differs from the rest. We observe this trend also in Figure B.1a with a global minima being identified towards the tail end of the time series.

The number of surgeries it requires to reach proficiency before the start of the pandemic is also akin to the inflection point at 93 with CUSUM in Figure B.1d. However, with BOCD analysis a change point from the proficient into the expert phase takes place at 360 in Figure B.3. The BOCD is therefore an improvement on the CUSUM when applied to the post-Covid - 19 data. This is because in Figure B.1g no learning curve was found due to an inflection point being found only for a global minima. Recall that this does not translate as being a transition into a more proficient stage but rather a deterioration in performance.

BOCD of cutting time

With change point probability being set to 0.4 for the total bone cutting time the BOCD algorithm identified six change points taking place at 55, 112, 176, 233, 321 and 372 in Figure B.4. The algorithm is very sensitive for this surgical step and more often than not detects a change point based on abnormally large time recorded as opposed to the probability of data being different in amongst those phases.

We observe there is insignificant difference in the average or standard deviation of the phase time. Similar to the bone sawing, bone cutting is a surgical step performed in conventional surgery also. This would

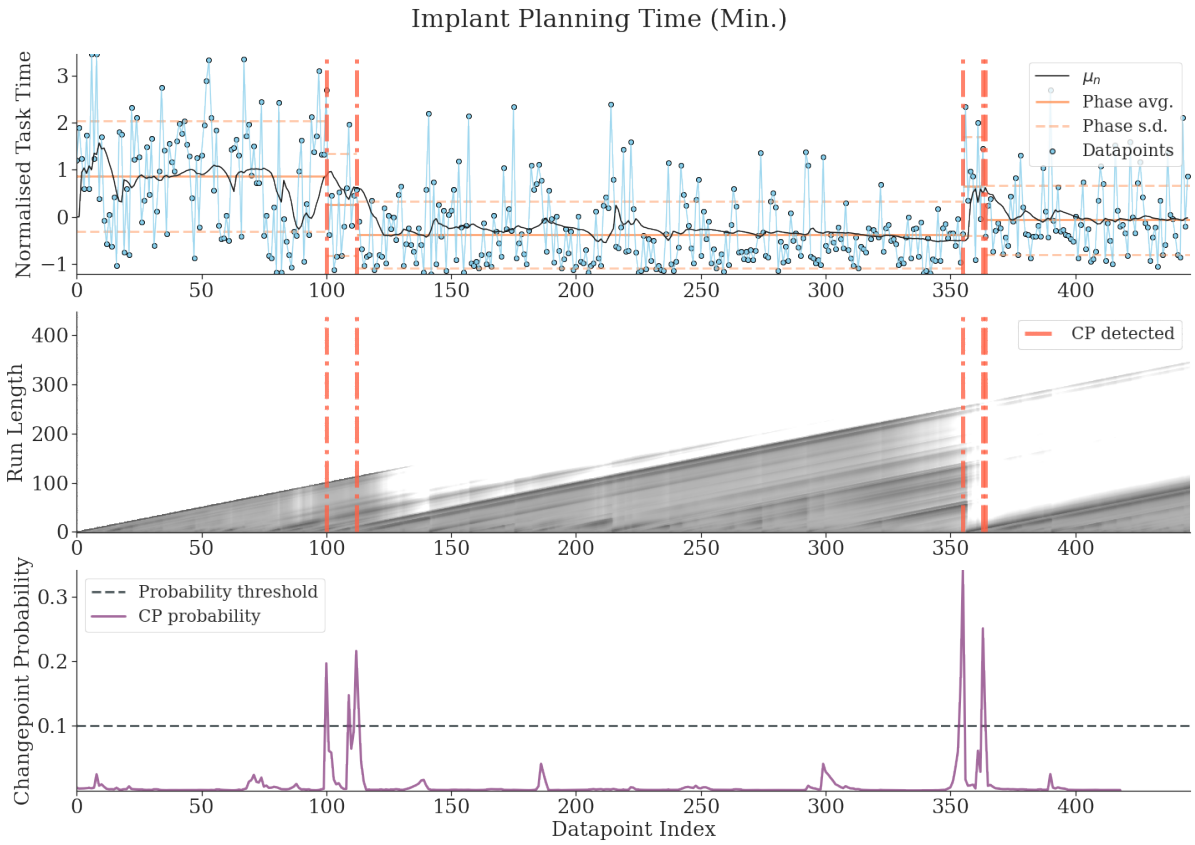


Figure B.3: BOCD for the implant planning time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1, \beta_0 = 1, \kappa_0 = 1, \mu_0 = 0$.

therefore provide an edge to experienced surgeons when working with the Mako RAS system and could explain why we witness so little improvement amongst the learning phases. The cut order is also influential on the flow of the procedure, with the first 71 surgeries having a different cut order to the rest of the data. The performance for the bone cutting time up to the 71 surgery mark is therefore incomparable with the rest of the data in a fair and accurate manner.

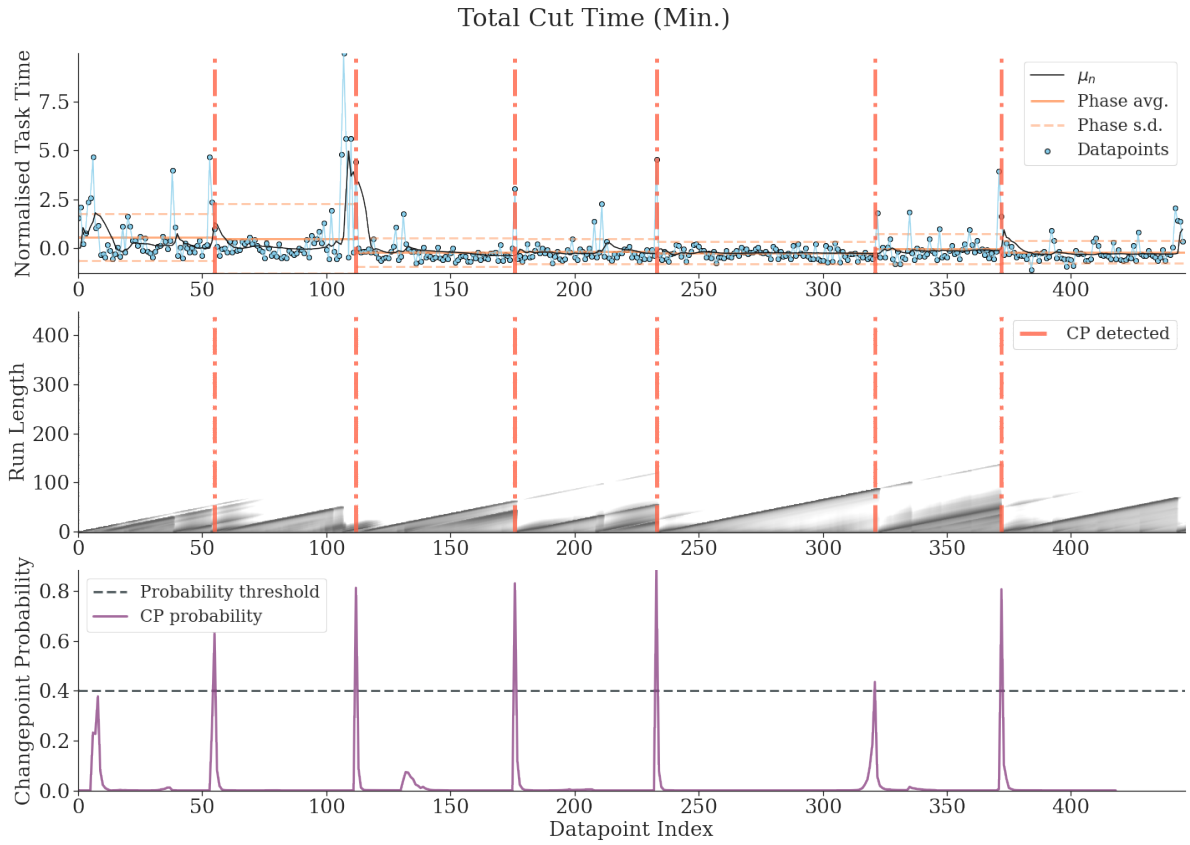


Figure B.4: BOCD for the total cutting time in minutes of Surgeon 1. Predictive probability π_t^r is modelled with parameters $\alpha_0 = 1$, $\beta_0 = 1$, $\kappa_0 = 1$, $\mu_0 = 0$.

This page intentionally left blank.