May 2023

Change Point Detection for a Process Having Several Regimes

Oliver Gerd Meister
University of Wisconsin-Milwaukee

CHANGE POINT DETECTION FOR A PROCESS HAVING SEVERAL REGIMES

by

Oliver Meister

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Mathematics

at

The University of Wisconsin-Milwaukee
May 2023

ABSTRACT

CHANGE POINT DETECTION FOR A PROCESS HAVING SEVERAL REGIMES

by

Oliver Meister

The University of Wisconsin-Milwaukee, 2023
Under the Supervision of Professor Richard Stockbridge

In this dissertation, possible methods for multiple change point detection on Markov chain processes are studied. Related works for offline and online change point detection are discussed and their applicability on sequential multiple change point detection for several regimes is evaluated. We develop a method for a multiple change point detection for a process having three regimes. Its efficiency is then evaluated on simulated Markov chain data by looking into different scenarios such as processes that significantly differ between each other or probability distributions that are slightly similar. This approach is then applied on Covid-19 hospital data. Therefore, the data is modeled into three different Markov chain processes and then used to successfully apply the derived change point detection method. In the end, the possible enhancements and its applications in other real world examples are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to Professor Richard Stockbridge for his invaluable support and advice throughout my whole time as a graduate student. Thanks should also go to Professor Chao Zhu and Professor David Spade for their feedback and patience in our weekly research meetings.

I am also grateful for my family and friends, especially my mom, grandpa and brother, giving me the moral support I needed in stressful time. Their belief in me during my whole academic career kept me motivated and made me believe in myself.

Lastly, I would like to say thank you to Brandon Schultz. His unconditional emotional support and patience can not be underestimated.

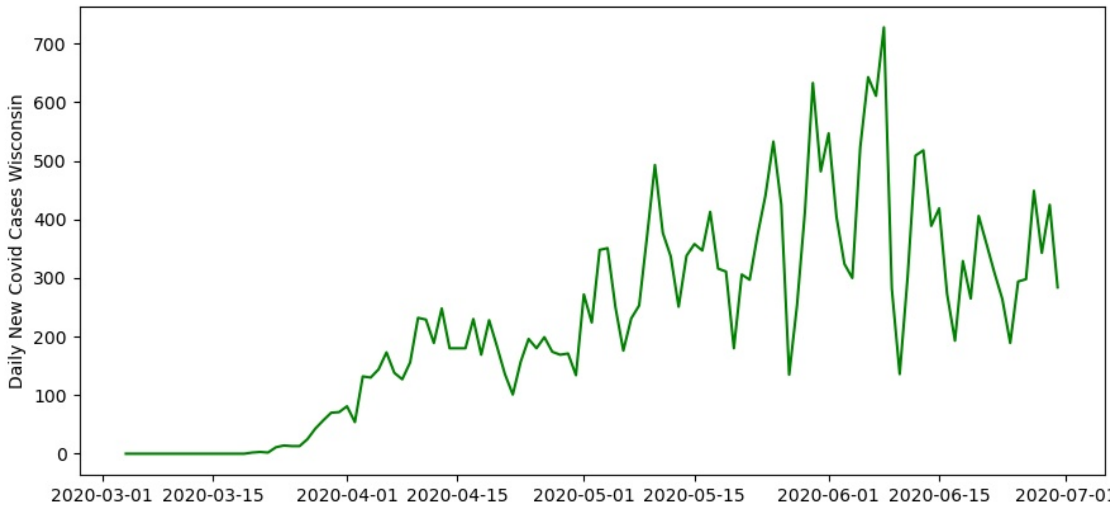
Oliver Meister

1.

Introduction

1.1 Motivation

In the beginning of the year 2020, SaRS-CoV-2 or better known as Covid-19 was a new virus that started spreading from Wuhan, a city in the People's Republic of China. The United States government was first hesitant to react when the initial cases of the virus were confirmed within the country. Scientists and politicians were indecisive when it was time to proclaim safety measurements to prevent the spread of the virus. The next months were shaped by arguments about the best time establishing new steps to contain the spread of the virus. Of course, these decisions were impacted by the constant change of daily Covid-19 cases. An increasing number of cases usually represented that the current measurements were not sufficient or that previous measurements were lifted too early. A decreasing number, on the other hand, represented that the measurements set in place were in fact working at the given time. One of the main questions to be addressed is how do we decide if the Covid-19 measurements are actually successful? If we take the Covid-19 cases and their change, we would say that mandates such as wearing a mask were successful if it led to a decline in cases. On the other hand, lifting the mask mandate would be considered too soon, if the cases began rising significantly after. The idea can be explained through change point detection theory. If we could see a change in cases right after masks became mandatory, we can say that it had an impact on the number of cases. Our goal is to detect these change points to evaluate how different decisions made during the pandemic impacted the cases. In addition, change points can help to decide when it is time to establish more measurements.



If we look at the first half of 2020, we can detect several change points. The first change point appears around the first cases starting March 10th. The number of new infections is increasing exponentially until April 10th, when the infection rate started to decay. This observation coincides with the lockdown that was issued on March 25th. Considering the 14 days delay that is caused by the incubation period, we are able to see that the lockdown helped to stop the exponential growth. Looking closer at the graph we can see more change points. The cases are starting to increase at the end of April which is when the stay at home order was lifted. Our mathematical objectives are to find a model that describes the evolution of infections as well as to develop an approach for a sequential multiple change point detection.

1.2 Model

Our change point detection will be based on Covid-19 cases. Therefore, it is essential to use a model that describes the evolution of infections. Epidemics or pandemics are often modeled in SIR (Susceptible-Infected-Recovered) or SEIR (Susceptible-Infected-Exposed-Recovered) processes which were established in the early 20th century with important works conducted by Ross Ross (1916), Ross and Hudson Ross and Hudson (1917), Kermack and McKendrick Kermack and McKendrick (1927) and Kendall Kendall (1956). SIR models are often run with ordinary differential equations, but are also used for stochastic frameworks to predict disease spread, total number of infections, or a duration of the pandemic.

Depending on the model being used, there are 3 to 4 stages:

(S) The susceptible class: Individuals that are able to become infected.

(E) The exposed class: Individuals that got exposed to the disease.

(I) The infected class: Individuals that became infected and are able to infect others from the susceptible class.

(R) The recovered class: Individuals that recovered from the infection. They can be considered immune or transmit to the susceptible class again.

At the beginning of our work we are using a simplified SIR Model where the birth and death probabilities are omitted. In addition, recovered individuals are not able to get infected again and therefore considered immune to the virus. Our goal is to use more advanced models, once we have established a change point detection method.

For the underlying prediction model we consider a discrete Markov Chain model with state vector $Y_t = \{S_t, I_t, R_t\}$. Define Y_i as the observation at time i whose distributions depend on the regime z_t.

$$Y_0, \cdots, Y_{\theta_0-1}, Y_{\theta_0}, \cdots, Y_{\theta_1-1}, Y_{\theta_1}, \cdots, Y_{\theta_m-1}, Y_{\theta_m}, \cdots, Y_n, \cdots$$

The change points $\theta_0, \dots, \theta_m$ initiate a possible change in the regime and therefore a change of the random variables' distribution.

2.

Literature Review

Before developing a method for finding change points in an online setup, it is necessary to consider existing methods. The studied methods can be divided into two categories: retrospective change point detection and online change point detection. While the retrospective approach analyzes the data after the observation, the online change point detection is looking for a sequential approach.

Retrospective methods are discussed in several papers such as “Detecting change-points in Markov chains” by Alan M. Polanksy (2007). Here, a likelihood theory is developed for a known number of changes, while AIC and BIC measures are used for the case of an unknown number of changes.Polansky (2007) Other methods of offline detection are mentioned in “Multiple Change-Point Detection via a Screening and Ranking Algorithm” by Ningo Hao, Yue Selena Niu and Heping Zhang (2013). The authors consider a sequence of random variables whose underlying step function has an unknown number of steps and unknown change points. The objective of their paper is to characterize the theoretical properties of the Screening and Ranking algorithm, as well as to develop a false discovery rate approach to the multiple change point problems.Hao et al. (2013)

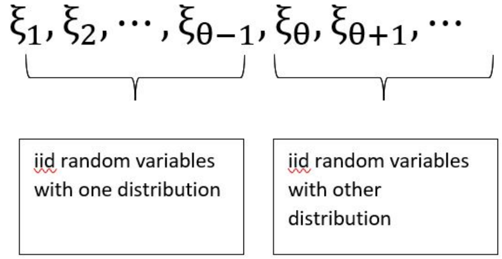
Finding a single change point in a sequential manner is mentioned in papers such as “Sequential Changepoint Detection in Quality Control and Dynamical System” by Tze Leung

Lais. A common approach is the CUSUM scheme where a cumulative sum stopping rule is defined which is optimal in the sense that the worst case delay is minimized. In addition the CUSUM scheme minimzes the conditional expected delay.Lai (1995) Other papers, describing the CUSUM methods, are written by Koepcke, Ashida and Kretzberg (2016)L. Koepcke (2016) or Yakir (1994)Yakir (1994). In the book "Optimal Stopping Rules" (2008) Shiryaev derives a method to find a single change point by converting change point detection into a optimal stopping problem.Shiryaev (2008) The author uses a Bayesian approach which is later modified by other author's such as Yakir for a Markov Chain model.

In the following sections we are first going to explain Shiryaev's approach followed by Yakir's adaption to a Markov Chain model completed by the proof which is omitted in Yakir's paper.

2.1 Shiryaev 2008 Optimal Stopping Rules

In Chapter 4.3 "The problem of disruption (discrete time)" of the book "Optimal Stopping rules"(2008), Shiryaev uses a Bayes formulation to convert change point detection into a optimal stopping time problem. On a measure space $(\Omega, \mathscr{F})$, we are given i.i.d. random variables $\xi_1, \xi_2, \dots$ and a probability measure P^π such that $P^\pi\{\theta = 0\} = \pi$ and $P^\pi\{\theta = n\} = (1-\pi)(1-p)^{n-1}p$ for $n \geq 1$ where the random variable θ is defined as the change point; and p and π are unknown constants with $0 < p < 1$ and $\pi \in [0,1]$. If $\theta = i$, the random variables $\xi_1, \dots, \xi_{i-1}, \xi_i, \dots$ are mutually independent. Furthermore, $\xi_1, \dots, \xi_{i-1}$ are identically distributed by the density $p_0(x)$ and $\xi_i, \xi_{i+1}, \dots$ are identically distributed by $p_1(x)$.



The change point θ is assumed to be geometrically distributed. The stopping time, the time at which the "alarm" signals a change point based on the observed process, is called τ. The risk associated with τ is defined as $\rho^{\pi}(\tau) = P^{\pi}\{\tau < \theta\} + c\mathbb{E}[\max\{\tau - \theta, 0\}]$, where the first term is interpreted as the probability of a false alarm or early detection, and the second term is the average delay of detecting the change point correctly ($\tau \geq \theta$). In order to understand Theorem 2.1 we need to define a π-Bayes time.

For a given $\pi \in [0, 1]$ a stopping time is called Bayes time if $\rho^{\pi}(\tau) = \inf \rho^{\pi}(\tau)$ where the infimum is taken over the class of all stopping times.

Theorem 2.1 now gives us a formula that detects a change point in a way that the risk $\rho^{\pi}(\tau)$ is minimized:

Theorem 2.1. *Let $c > 0, p > 0$, and let $\pi_n^{\pi} = P^{\pi}\{\theta \leq n | \mathscr{F}_n^{\xi}\}$ be a posterior probability of disruption occuring before time n; $\pi_0^{\pi} = \pi$. Then there exists a constant A^* such that the τ_{π}^* is a Bayes time with*

$$\tau_{\pi}^* = \inf\{n \geq 0 : \pi_n^{\pi} \geq A^*\}$$

where

$$\mathscr{F}_0^{\xi} = \{\emptyset, \Omega\}$$

$$\mathscr{F}_n^{\xi} = \sigma\{\xi_1, \xi_2, \ldots, \xi_n\}.$$

2.2 Yakir 1994, Optimal Detection of a Change in Distribution when the Observations Form a Markov Chain with a Finite State Space

In Yakir's paper"Optimal Detection Of A Change in Distribution When The Observations Form A Markov Chain With A Finite State Space", the assumption of independence of the observations is lifted by considering a Markov chain process with a finite state space. One approach is the Bayes formulation as explained in Shiryaev's book for independent random variables.

Let $(\Omega, \mathscr{F})$ be a measure space on which we are given the random variables $X_0, X_1, \ldots$ and the change point θ. Furthermore, the author assumes a probability measure $P^{\pi,x}$ such that:

$$P^{\pi,x}(X_0 = x) = 1,$$

$$P^{\pi,x}(\theta = 1) = \pi$$

and:

$$P^{\pi,x}(\theta = k) = (1-\pi)p(1-p)^{k-2}, \text{ for } k \geq 2:$$

where p, π and x are known constants with $0 < p < 1$, $0 \leq \pi \leq 1$ and x one of the possible state of the process $(x \in X)$. The parameter θ is the random change point which is geometrically distributed.

The distribution of our observations depends on the random change point θ. Assuming that our change point $\theta = k$ we get the distribution of the random process $X_1, X_2, \ldots$:

$$P^{\pi,x}(X_1 = x_1, X_2 = x_2, \ldots | \theta = k) = \begin{cases} \prod_{i=1}^{k-1} a(x_i|x_{i-1}) \prod_{i=k}^{n} b(x_i|x_{i-1}), & \text{if } k \leq n \\ \prod_{i=1}^{n} a(x_i|x_{i-1}), & \text{if } k > n. \end{cases}$$

The matrices $A = (a(i|j))$ and $B = (b(i|j))$ are two transition probability matrices for the Markov process. Before the change point θ the distribution of the Markov chain has the transition matrix A. If the change point occurred, the transition matrix changes to B. Yakir's next step is to define a stopping rule that is based on a risk function ρ. Therefore, the stopping time N is adapted to the system of σ-algebras $\{\mathscr{F}_n\}_{n=0}^{\infty}$ which can be interpreted as the knowledge obtained by making the observations $x_1, \ldots, x_n$. The risk function is based on the definition of Shiryaev's paper:

$$\rho(N, \theta) = P^{(\pi,x)}(N < \theta - 1) + c\mathbb{E}^{(\pi,x)}(N - \theta + 1)^{+}$$

where $c > 0$ is a fixed constant. The difference between Shiryaev's and Yakir's defined risk function is that Yakir waives the cost for signaling an alarm one time step before or after the actual change point. Yakir also uses the Bayes' time to characterize the stopping rule for minimizing the risk function (Theorem 1):

Theorem 2.2. *Let $p > 0, c > 0$ then there exist a function $\delta(X_n)$ such that*

$$N^* = \inf\{n \geq 0 : \pi_n^{(\pi,x)} \geq \delta(X_n)\},$$

is the (π, x)-Bayes' rule. Moreover, $\delta(\cdot)$ does not depend on π or x.

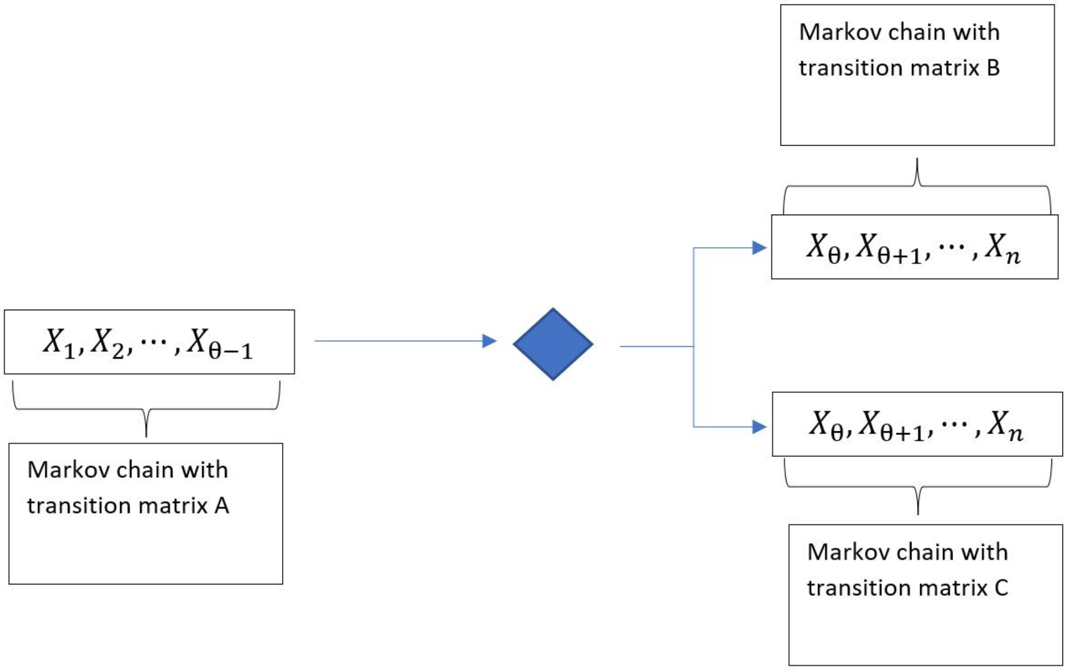
3.

Motivation

After comprehending Shiryaev's and Yakir's approach in finding a single change point in an online matter, we wish to modify Yakir's method to a model with multiple alternative regimes. We are choosing Yakir because his detection is based on a Markov chain model. In the following sections we will explain the different models that we define a change point detection for.

3.1 Change Point Detection With Three Regimes

The first modification on Yakir's approach would be to extend the number of regimes to three. There is an initial regime A where our Markov random variables have a transition matrix A. After a change point occurred there are two possible regimes. Therefore, the Markov random variables $X_\theta, X_{\theta+1}, \ldots, X_n$ can either have transition Matrix B or C.



In order to modify the change point detection, we have to analyze how the different regimes are affecting the posterior probability, which is necessary to find the Bayes' time. Furthermore, we will extend the risk function. Next to having a penalty for an early detection and a detection delay, we will have to consider the risk for choosing the wrong regime.

Let the change points to regime B and C be defined as $\theta_B \sim$ geom (λ_B) and $\theta_C \sim$ geom (λ_C) where θ_B and θ_C are independent random variables. A change point θ occurred if the process either shifted to B or C. Therefore, the new change point is defined as $\theta = \min(\theta_B, \theta_C)$. For the case that the change points occur at the same time, we define that the regime will switch to B.

The following lemmas will helps us define a posterior probability and extending our risk function.

Lemma 3.1. *Let $\theta_B \sim$ geom (λ_B) and $\theta_C \sim$ geom (λ_C) be independent random variables. Define $\theta = \min(\theta_B, \theta_C)$. Then $\theta \sim geom(\lambda_B + \lambda_C - (\lambda_B \cdot \lambda_C))$.*

Proof.

$$P(\theta_B \geq k) = (1 - \lambda_B)^{k-1}$$

$$P(\theta_C \geq k) = (1 - \lambda_C)^{k-1}$$

$$P(\min(\theta_B, \theta_C) \geq k) = P(\theta_B \geq k, \theta_C \geq k)$$

$$= P(\theta_B \geq k)P(\theta_C \geq k) = (1 - \lambda_B)^{k-1}(1 - \lambda_C)^{k-1}$$

$$P(\min(\theta_B, \theta_C) = k) = P(\min(\theta_B, \theta_C) \geq k) - P(\min(\theta_B, \theta_C) \geq k + 1)$$

$$= (1 - \lambda_B)^{k-1}(1 - \lambda_C)^{k-1} - (1 - \lambda_B)^k(1 - \lambda_C)^k$$

$$= (\lambda_B + \lambda_C - \lambda_B\lambda_C)((1 - \lambda_B)(1 - \lambda_C))^{k-1}$$

$$= (\lambda_B + \lambda_C - \lambda_B\lambda_C)(1 - (\lambda_B + \lambda_C - \lambda_B\lambda_C))^{k-1}$$

□

Lemma 3.2. *Let $\theta_B \sim$ geom (λ_B) and $\theta_C \sim$ geom (λ_C) be independent random variables. Then $P(\theta_B < \theta_C) = \frac{\lambda_B(1-\lambda_C)}{\lambda_B+\lambda_C-\lambda_B\lambda_C}$*

Proof.

$$
\begin{aligned}
P(\theta_B < \theta_C) &= \sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} (1-\lambda_B)^x (1-\lambda_C)^y \lambda_B \lambda_C \\
&= \lambda_B \lambda_C \sum_{x=0}^{\infty} (1-\lambda_B)^x \sum_{y=x+1}^{\infty} (1-\lambda_C)^y \\
&= \lambda_B \lambda_C \sum_{x=0}^{\infty} (1-\lambda_B)^x [\sum_{y=0}^{\infty} (1-\lambda_C)^y - \sum_{y=0}^{x} (1-\lambda_C)^y] \\
&= \lambda_B \lambda_C \sum_{x=0}^{\infty} (1-\lambda_B)^x \left[\frac{1}{1-(1-\lambda_C)} - \frac{1-(1-\lambda_C)^{x+1}}{1-(1-\lambda_C)}\right] \\
&= \lambda_B \lambda_C \sum_{x=0}^{\infty} (1-\lambda_B)^x \left[\frac{(1-\lambda_C)^{x+1}}{\lambda_C}\right] \\
&= \lambda_B (1-\lambda_C) \sum_{x=0}^{\infty} [(1-\lambda_B)(1-\lambda_C)]^x \\
&= \lambda_B (1-\lambda_C) \frac{1}{1-(1-\lambda_B)(1-\lambda_C)} \\
&= \frac{\lambda_B(1-\lambda_C)}{\lambda_B + \lambda_C - \lambda_B \lambda_C}
\end{aligned}
$$

□

Lemma 3.3. $P(\theta_B = \theta_C) = \frac{\lambda_B \lambda_C}{\lambda_B + \lambda_C + \lambda_B \lambda_C}$

Proof.

$$
\begin{aligned}
P(\theta_B = \theta_C) &= 1 - P(\lambda_B < \lambda_C) - P(\lambda_C < \lambda_B) \\
&= 1 - \frac{\lambda_B(1-\lambda_C)}{\lambda_B + \lambda_C - \lambda_B \lambda_C} - \frac{\lambda_C(1-\lambda_B)}{\lambda_B + \lambda_C - \lambda_B \lambda_C} \\
&= \frac{\lambda_B + \lambda_C - \lambda_B \lambda_C - \lambda_B + \lambda_B \lambda_C - \lambda_C + \lambda_B \lambda_C}{\lambda_B + \lambda_C - \lambda_B \lambda_C} \\
&= \frac{\lambda_B \lambda_C}{\lambda_B + \lambda_C - \lambda_B \lambda_C}
\end{aligned}
$$

□

The first step is to find the posterior probability $P(\theta < n|\mathscr{F})$ that a change point occurred before time n, π_n. We are extending the condition of the posterior probability by the regime that we are in. In our case, we have two cases: regime B and regime C. The new posterior probability π_n is defined as $P(\theta \leq n|\mathscr{F}, i)$ where i is the regime.

$$\pi_{n+1,i} = \frac{\pi_n p_i(\xi_{n+1}) + (1-\pi_n)\lambda_B(1-\lambda_C)p_B(\xi_{n+1}) + (1-\pi_n)(1-\lambda_B)\lambda_C p_C(\xi_{n+1})}{\text{numerator } + (1-\pi_n)(1-\lambda_B)(1-\lambda_C)p_0(\xi_{n+1})} + \frac{+(1-\pi_n)\lambda_B\lambda_C p_B(\xi_{n+1})}{\text{numerator } + (1-\pi_n)(1-\lambda_B)(1-\lambda_C)p_0(\xi_{n+1})}$$

$$\pi_{n+1,B} = P(\theta \leq n, \alpha = B|\mathscr{F}) = \frac{\pi_{n,B}p_B(\xi_{n+1}) + (1-\pi_{n,B})(1-\pi_{n,C})(\lambda_B+\lambda_C-\lambda_B\lambda_C)\frac{\lambda_B}{\lambda_B+\lambda_C-\lambda_B\lambda_C}p_B(\xi_{n+1})}{numerator + (1-\pi_{n,B})(1-\pi_{n,C})(1-\lambda_B+\lambda_C-\lambda_B\lambda_C)p_A(\xi_{n+1})}$$

and

$$\pi_{n+1,C} = P(\theta \leq n, \alpha = C|\mathscr{F}) = \frac{\pi_{n,C}p_C(\xi_{n+1}) + (1-\pi_{n,C})(1-\pi_{n,B})(\lambda_C+\lambda_B-\lambda_C\lambda_B)\frac{\lambda_C(1-\lambda_B)}{\lambda_C+\lambda_B-\lambda_C\lambda_B}p_C(\xi_{n+1})}{numerator + (1-\pi_{n,C})(1-\pi_{n,B})(1-\lambda_C+\lambda_B-\lambda_C\lambda_B)p_A(\xi_{n+1})}$$

4.

Change Point Detection For A Markov Chain Model With Three Regimes

In this chapter, we are adapting Yakir's approach, described in section 2.2 of a change point detection with one alternative regime, to a method that detects a change point for a Markov chain model with two alternative regimes.

4.1 Markov Chain Model With Three Regimes

Our goal is to derive a method that detects a change point with two alternative regimes. We again assume $X_1, X_2, \ldots$ are random variables governed by a Markov Chain distribution with

$$P(X_0 = x) = 1$$

$$P(x_i|x_{i-1}) = a(x_i|x_{i-1})$$

for $i = 1, 2, \ldots, \theta - 1$. After the change point, $\theta = \min\{\theta_B, \theta_C\}$ where $\theta_B \sim \text{geom}(\lambda_B)$ and $\theta_C \sim \text{geom}(\lambda_C)$ are i.i.d. random variables, we have the probability distribution

$$P(x_i|x_{i-1}) = b(x_i|x_{i-1})$$

if $\theta = \theta_B$ and

$$P(x_i|x_{i-1}) = c(x_i|x_{i-1})$$

if $\theta = \theta_C$

for $i = \theta, \theta + 1, \ldots$

Therefore we have:

$$P(\theta = 0) = P(\theta_B = 0) + P(\theta_C = 0, \theta_B > 0) = \pi_B + (1 - \pi_B) \cdot \pi_C = \pi_0 \ \in (0, 1)$$

with $\pi_B = P(\theta_B = 0)$ and $\pi_C = P(\theta_C = 0)$.

Furthermore,

$$P(\theta = n) = (1 - \pi_0)(1 - p)^{n-1} p, \quad n = 1, 2, \ldots$$

where $p = \lambda_B + \lambda_C - \lambda_B \cdot \lambda_C$.

The posterior probabilities are defined as follows

Definition 4.1. *Let α denote the current regime of the process, then the probability for a change point to regime α, occurring before or at time n conditioned by $\mathscr{F}_n$ is given by:*

$$\pi_{n,\alpha} = P(\theta \leq n, \alpha_n = \alpha|\mathscr{F}_n)$$

Therefore, for our model, we have

$$\pi_{n,B} := P(\theta \leq n, \alpha_n = B|\mathscr{F}_n)$$

and

$$\pi_{n,C} := P(\theta \leq n, \alpha_n = C|\mathscr{F}_n)$$

where α_n is the regime at time step n and

$$\mathscr{F}_0 = \{\emptyset, \Omega\}$$

$$\mathscr{F}_n = \sigma\{X_0, X_1, X_2, \ldots, X_n\}$$

4.2 Proof that the new posterior probability is a Markov random function

In order to apply Yakir's approach, we have to show that the posterior function for the change point is a Markov function. The definition of a Markov function can be found in Shiryaev.

Definition 4.2 (Markov Random Function Shiryaev (2008)). *The system $X = (x_t, \mathscr{F}_t, P)$,$t \in Z$ is said to be a (homogeneous nonterminating) Markov random function if (1.36) is satisfied and if for all s,t $\in Z, B \in \mathscr{B}$,*

$$P(X_{t+s} \in B|\mathscr{F}_t) = P(x_{t+s} \in B|x_t)$$

Definition 4.3 (Markov Process Shiryaev (2008)). *The system $X = (x_t, \mathscr{F}_t, P_x), t \in \mathscr{Z}, x \in E$, is said to be a (homogeneous, nonterminating) Markov process with values in a state space $(E, \mathscr{B})$ if the following conditions are satisfied:*

(1) For each $A \in \mathscr{F}, P_x(A)$ is a $\mathscr{B}$-measurable function for x;

(2) For all $x \in E, B \in \mathscr{B}, s, t \in \mathscr{Z}$,

$$P_x(x_{t+s} \in B|\mathscr{F}_t) = P_{x_t}(x_s \in B) \qquad (1.35)$$

(3) $P_x(x_0 = x) = 1, x \in E$;

(4) for each $t \in \mathscr{Z}$ and $\omega \in \Omega$ there will be a unique $\omega' \in \Omega$ such that

$$x_s(\omega') = x_{s+t}(\omega) \tag{1.36}$$

for all $s \in \mathscr{Z}$

We are now showing that the posterior functions are Markov functions

Lemma 4.4. *Assume that we are having the model described in section 4.1 and let $A_n = X_1, \ldots, X_n$. Then both $\pi_{n,B}$ and $\pi_{n,C}$ are Markov random functions.*

Proof. We are first showing that $\pi_{n,B}$ is a Markov random function.

The posterior function with a transition to regime B can be calculated as:

$$\begin{aligned}\pi_{n,B} = &\frac{\pi_B \prod_{i=1}^{n} b_{i-1,i} + (1-\pi_0)\sum_{k=2}^{n}\prod_{i=1}^{k-1} a_{i-1,i} q^{k-2}\lambda_B \prod_{i=k}^{n} b_{i-1,i}}{P(A)} \\ &+ \frac{+(1-\pi_0)\prod_{i=1}^{n} a_{i-1,i} q^{n+1-2}(\lambda_B)}{P(A)}\end{aligned} \tag{4.5}$$

with $p = \lambda_B + \lambda_C(1-\lambda_B)$, $q = 1-p$, $\pi_0 = \pi_B + (1-\pi_B)\pi_C$, $a_{i-1,i} = a(x_i|x_{i-1})$ and $b_{i-1,i} = b(x_i|x_{i-1})$, and

$$\begin{aligned}P(A) =&\pi_B \prod_{i=1}^{n} b_{i-1,i} + (1-\pi_0)\sum_{k=2}^{n}\prod_{i=1}^{k-1} a_{i-1,i} q^{k-2}\lambda_B \prod_{i=k}^{n} b_{i-1,i} + (1-\pi_0)\prod_{i=1}^{n} a_{i-1,i} q^{n+1-2}(\lambda_B) \\ &+ (1-\pi_B)\pi_C \prod_{i=1}^{n} c_{i-1,i} + (1-\pi_0)\sum_{k=2}^{n}\prod_{i=1}^{k-1} a_{i-1,i} q^{k-2}\lambda_C(1-\lambda_B)\prod_{i=k}^{n} c_{i-1,i} \\ &+ (1-\pi_0)\prod_{i=1}^{n} a_{i-1,i} q^{n+1-2}(\lambda_C)(1-\lambda_B) \\ &+ (1-\pi_0)\sum_{k=n+2}^{\infty}\prod_{i=1}^{n} a_{i-1,i} q^{k-2} p.\end{aligned} \tag{4.6}$$

The first term in the numerator describes the case where a change point happens at time step 0. The second term is calculating the probability that a change point occurs between time step 1 and n. The last term shows the probability that a change point happens at time $n+1$. In that case no observation is governed by B yet.

We can rewrite the first two terms of the numerator as follows:

$$\begin{aligned}
&\pi_B \prod_{i=1}^{n} b_{i-1,i} + (1-\pi_0) \sum_{k=2}^{n} \prod_{i=1}^{k-1} a_{i-1,i} q^{k-2} \lambda_B \prod_{i=k}^{n} b_{i-1,i} \\
&= \pi_B \prod_{i=1}^{n-1} b_{i-1,i} \times b_{n-1,n} + (1-\pi_0) \sum_{k=2}^{n-1} \prod_{i=1}^{k-1} a_{i-1,i} q^{k-2} \lambda_B \prod_{i=k}^{n-1} b_{i-1,i} \times b_{n-1,n} \\
&\quad + (1-\pi_0) \prod_{i=1}^{n-1} a_{i-1,i} q^{n-1} \lambda_B \times b_{n-1,n} \\
&= \left[\pi_B \prod_{i=1}^{n-1} b_{i-1,i} + (1-\pi_0) \sum_{k=2}^{n-1} \prod_{i=1}^{k-1} a_{i-1,i} q^{k-2} \lambda_B \prod_{i=k}^{n-1} b_{i-1,i} + (1-\pi_0) \prod_{i=1}^{n-1} a_{i-1,i} q^{n-1} \lambda_B \right] \times b_{n-1,n} \\
&= \pi_{n-1,B}\, b_{n-1,n}\, P(A_n).
\end{aligned} \tag{4.7}$$

The same can be shown for the numerator of $\pi_{n,C}$.

Next we consider the last term in the denominator. We use the fact that for the geometric series, it holds that $\sum_{k=n+2}^{\infty} q^{k-2} = q^n \frac{1}{1-q} = \frac{q^n}{p}$.

$$\begin{aligned}
&(1-\pi_0) \sum_{k=n+2}^{\infty} \prod_{i=1}^{n} a_{i-1,i} q^{k-2} p \\
&= (1-\pi_0) \prod_{i=1}^{n} a_{i-1,i} \frac{q^n}{p} p \\
&= (1-\pi_0) \prod_{i=1}^{n} a_{i-1,i} q^n \\
&= (1-\pi_0) \prod_{i=1}^{n-1} a_{i-1,i} q^{n-1}\, a_{n-1,n}\, q.
\end{aligned} \tag{4.8}$$

Furthermore, $(1-\pi_0)\prod_{i=1}^{n-1} a_{i-1,i}q^{n-1} = P(\theta - 1 > n-1, \mathscr{F}_{n-1})$. If we compare the last term of the numerator with these results, we can see that

$$\begin{aligned}
&(1-\pi_0)\prod_{i=1}^{n} a_{i-1,i}q^{n+1-2}(\lambda_B) \\
&= (1-\pi_0)\prod_{i=1}^{n-1} a_{i-1,i}q^{n-1}(\lambda_B)\ a_{n-1,n} \\
&= P(\theta - 1 > n-1, \mathscr{F}_{n-1})\ \lambda_B\ a_{n-1,n}.
\end{aligned} \tag{4.9}$$

These results show that we can rewrite our posterior probability as follows:

$$\frac{\pi_{n-1,B}\ b_{n-1,n}\ P(A_n) + P(\theta > n-1, \mathscr{F}_{n-1})\ \lambda_B\ a_{n-1,n}}{numerator_B + numerator_C + P(\theta > n-1, \mathscr{F}_{n-1})\ q\ a_{n-1,n}} \tag{4.10}$$

with

$$numerator_B = \pi_{n-1,B}\ b_{n-1,n} + (1 - (\pi_{n-1,B} + \pi_{n-1,C}))\ \lambda_B\ a_{n-1,n}$$

and

$$numerator_C = \pi_{n-1,C}\ c_{n-1,n} + (1 - (\pi_{n-1,B} + \pi_{n-1,C})\ (\lambda_C(1-\lambda_B))\ a_{n-1,n}.$$

Dividing each term by $P(A_n)$ gives us the following result:

$$\begin{aligned}
\pi_{n,B} =& \frac{\pi_{n-1,B}\ b_{n-1,n} + P(\theta > n-1 | \mathscr{F}_{n-1})\ \lambda_B\ a_{n-1,n}}{numerator_B + numerator_C + P(\theta > n-1 | \mathscr{F}_{n-1})\ q\ a_{n-1,n}} \\
=& \frac{\pi_{n-1,B}\ b_{n-1,n} + (1 - (\pi_{n-1,B} + \pi_{n-1,C}))\ \lambda_B\ a_{n-1,n}}{numerator_B + numerator_C + \ (1 - (\pi_{n-1,B} + \pi_{n-1,C}))\ q\ a_{n-1,n}}.
\end{aligned} \tag{4.11}$$

In the end $numerator_C = \pi_{n-1,C}\ c_{n-1,n} + (1 - (\pi_{n-1,B} + \pi_{n-1,C}))\ \lambda_C(1-\lambda_B)\ a_{n-1,n}$. The last equation is true because $P(\theta - 1 \leq n | \mathscr{F}_n) = P(\theta - 1 \leq n, \alpha = B | \mathscr{F}_n) + P(\theta - 1 \leq n, \alpha = C | \mathscr{F}_n)$.

Furthermore, the last term shows that the posterior function $\pi_{n,B}$ is a Markov function.

We are getting a similar result for the posterior probability with regard to regime C:

$$\pi_{n,C} = \frac{\pi_{n-1,C}\; c_{n-1,n} + (1 - (\pi_{n-1,B} + \pi_{n-1,C})\; (\lambda_C(1 - \lambda_B))\; a_{n-1,n}}{numerator_B + numerator_C + (1 - (\pi_{n-1,B} + \pi_{n-1,C})\; q\; a_{n-1,n}}$$

and therefore, π_n is also a Markov function. □

4.3 Transforming the Change Point Detection Into an Optimal Stopping Time Problem

Similar to Shiryaev, we are defining a risk for causing an alarm at time τ. Consider the risk function $\rho^\pi(\tau) = P\{\tau < \theta\} + c\mathbb{E}(\tau - \theta)^+$. The first term describes an early detection and the second term an expected detection delay.

We are interested in writing the risk function into a formula depending on π_n.

Lemma 4.12. *For each $\tau \in \mathfrak{M}[F]$, we note that $P^\pi\{\tau < \theta\} = \mathbb{E}^\pi[1 - \pi^\pi_\tau]$ with $\mathfrak{M}[F]$ being the class of all stopping time with respect to the system $\mathscr{F}_n$.*

Proof.

$$
\begin{aligned}
\mathbb{E}[1-\pi_\tau] &= 1-\mathbb{E}[\pi_\tau] \\
&= 1-\sum_n \mathbb{E}[\pi_n I_{(\tau=n)}] \\
&= 1-\sum_n \mathbb{E}[1_{(\tau=n)}\mathbb{E}[I_{(\theta\le n)}|\mathscr{F}_n]] \\
&= 1-\sum_n E[E[1_{(\tau=n)}1_{(\theta\le n)}|\mathscr{F}_n]] \\
&= 1-\sum_n \mathbb{E}[1_{(\theta\le\tau,\tau=n)}] \\
&= 1-\mathbb{E}[\sum_n 1_{(\theta\le\tau,\tau=n)}] \\
&= 1-\mathbb{E}[1_{(\theta\le\tau)}] \\
&= P(\theta>\tau)
\end{aligned}
\tag{4.13}
$$

□

In addition, we have to derive a formula for the second term $\mathbb{E}(\tau-\theta)^+$. The proof is similar to Shiryaev Shiryaev (2008). We show that for each $n\ge 0$:

$$
\begin{aligned}
\mathbb{E}^\pi[\max(n-\theta,0)|\mathscr{F}_n] &= \sum_{k=0}^{n}(n-k)P^\pi\{\theta=k|\mathscr{F}_n\} \\
&= \sum_{k=0}^{n-1} P^\pi\{\theta\le k|\mathscr{F}_n\} \\
&= \sum_{k=0}^{n-1}[P^\pi\{\theta\le k|\mathscr{F}_n\}-P^\pi\{\theta\le k+1|\mathscr{F}_k\}]+\sum_{k=0}^{n-1}P^\pi\{\theta\le k|\mathscr{F}_k\} \\
&= \sum_{k=0}^{n-1}[P^\pi\{\theta\le k|\mathscr{F}_n\}-P^\pi\{\theta\le k|\mathscr{F}_k\}]+\sum_{k=0}^{n-1}\pi_k^\pi.
\end{aligned}
\tag{4.14}
$$

We define the sum as

$$\begin{aligned}\psi_n^\pi &= \sum_{k=0}^{n} \left[P^\pi\{\theta \le k|\mathscr{F}_n\} - P^\pi\{\theta \le k|\mathscr{F}_k\}\right] \\ &= -\sum_{k=0}^{n} \left[P^\pi\{\theta \ge k|\mathscr{F}_n\} - P^\pi\{\theta \ge k|\mathscr{F}_k\}\right].\end{aligned} \tag{4.15}$$

The sequence $\{\psi_n^\pi, \mathscr{F}_n, P^\pi\}$, $n \ge 0$ then forms a martingale for each $\pi \in [0, 1]$:

$$\begin{aligned}\mathbb{E}[\psi_{n+1}^\pi|\mathscr{F}_n] &= \mathbb{E}[\sum_{k=0}^{n+1}[P^\pi\{\theta \le k|\mathscr{F}_{n+1}\} - P^\pi\{\theta \le k|\mathscr{F}_k\}]|\mathscr{F}_n] \\ &= \mathbb{E}[\sum_{k=0}^{n}[P^\pi\{\theta \le k|\mathscr{F}_{n+1}\} - P^\pi\{\theta \le k|\mathscr{F}_k\}]|\mathscr{F}_n] \\ &= \mathbb{E}[\sum_{k=0}^{n}[P^\pi\{\theta \le k|\mathscr{F}_n\} - P^\pi\{\theta \le k+1|\mathscr{F}_k\}]] \\ &= \psi_n^\pi.\end{aligned} \tag{4.16}$$

We now use the Optional Sampling Theorem (Theorem 1.12, Shiryaev 2008) to show that for any $\tau \in \mathscr{M}[F]$

$$\mathbb{E}[\psi_\tau^\pi] = \mathbb{E}[\psi_0^\pi] = 0.$$

Hence for $\tau \in \mathscr{M}[F]$, we get the risk:

$$\begin{aligned}\rho^\pi(\tau) &= P^\pi\{\tau < \theta\} + c\mathbb{E}[\max\{\tau - \theta, 0\}] \\ &= \mathbb{E}[(1-\pi_\tau) + c\sum_{k=0}^{\tau-1} \pi_k^\pi + c\psi_\tau^\pi] \\ &= \mathbb{E}[(1-\pi_\tau^\pi) + c\sum_{k=0}^{\tau-1} \pi_k^\pi].\end{aligned} \tag{4.17}$$

Hence, for the Bayes time $\rho^\pi = \inf_{\tau\in\mathscr{M}[F]} \rho^\pi(\tau)$,

we find that $\rho^\pi = \inf_{\tau\in\mathscr{M}[F^\xi]} \mathbb{E}[(1-\pi_\tau) + c\sum_{k=0}^{\tau-1} \pi_k^\pi]$.

$\Pi^\pi = (\pi_n^\pi, \mathscr{F}_n^\xi, P^\pi), n \ge 0$, forms a sub-martingale: $\mathbb{E}[\pi_{n+1}^\pi|\mathscr{F}_n] \ge \pi_n^\pi \ \ (P^\pi - a.s.)$.

4.4 Optimal Stopping Time Without Penalty For Choosing The Wrong Regime

The family of Markov random functions $\{\Pi^{\pi,0\leq\pi\leq1}\}$ can be associated with a Markov process in discrete time $\Pi = (\pi_n, \mathscr{F}, P_\pi), n \geq 0$, having the same transition probabilities as each Markov random function $\Pi^\pi, \pi \in [0,1]$.Shiryaev (2008)

In Chapter 2.14, Shiryaev develops a method to find the optimal stopping time for a Markov sequence $X = (X_n, \mathscr{F}_n, P_x)$. The stopping time is optimal in sense of maximizing the gain for stopping at time n. The gain function is defined by

$$G(n, x_0, \ldots x_n) = \alpha^n g(x_n) - \sum_{s=0}^{n-1} \alpha^s c(x_s) \tag{4.18}$$

for $n \geq 0$ and $G(0, x_0) = g(x_0)$ where $g(x_n)$ is interpreted as the gain for stopping at time n while $c(x)$ are the costs for the opportunity of making the next observation. The variable α is the discount factor which we can disregard in our case.Shiryaev (2008)

The goal is to maximize the payoff

$$s(x) = \sup \mathbb{E}\{\alpha^\tau g(x_\tau) - \sum_{s=0}^{\tau-1} \alpha^s c(x_s)\} \tag{4.19}$$

where the supremum is taking over the stopping time classes $\mathscr{M}_{(\alpha.c)}$.

Theorem 4.20 (Shiryaev (2008))**.** *Let the functions $g(x)$ and $c(x)$ satisfy* (4.18), $0 < \alpha \leq 1$. *Then:*

(1) The payoff $s(x)$ is the smallest (α,c)- excessive majorant of the function $g(x)$;

(2) $s(x) = \max\{g(x), \alpha, Ts(x) - c(x)\}$

(3) $s(x) = \lim_{N\to\infty} Q^N_{\{\alpha,c\}} g(x)$ *where* $Q^N_{\{\alpha,c\}}$ *is the* N*th power of the operator*

(4) For any $\epsilon > 0$ *the time* $\tau_\epsilon = \inf\{n \geq 0 : \alpha^n s(x_n) \leq \alpha^n g(x_n) + \epsilon\}$ *is an* ϵ*-optimal stopping time for the class* $\mathscr{M}_{(\alpha,c)}$

(5) If $P_x(\tau_0 < \infty) = 1, x \in E$, *the time* τ_0 *will be an optimal stopping time for the class* $\mathscr{M}_{(\alpha,c)}$

(6) If $P_x\{\sum_{s=0}^{\infty} \alpha^s c(x_s = \infty)\} = 1, x \in E$ *then* $P_x\{\tau_0 < \infty\} = 1$ *and the time* τ_0 *is an optimal stopping time in the class* $\mathscr{M}_{(\alpha,c)}$

We derive the following method to find the optimal stopping time.

Theorem 4.21. *The optimal stopping time for the model described in Section 4.1 is*

$$\tau_0 = \inf\{n \geq 0 : \rho(\pi_n) = 1 - \pi_n\} = \inf\{n \geq 0 : \pi_n \geq A^*\}$$

with $\pi_n = \pi_{n,B} + \pi_{n,C}$.

Proof. Now consider the risk function

$$R(\pi_n, \tau) = 1 - \pi_\tau + c\sum_{k=0}^{\tau-1} \pi_k \tag{4.22}$$

With the results of Theorem 4.20, we only need to find the optimal stopping time for the payoff to find the π-Bayes time τ^*_π. However, we minimize our risk function. Therefore, we multiply $R(\pi_n, \tau)$ by -1:

$$-R(\pi_n, \tau) = -(1 - \pi_\tau) - c \sum_{k=0}^{\tau-1} \pi_k$$

We set $g(\pi) = -(1 - \pi)$, the probability of a early detection, as the gain function, and $c(\pi) = c \sum_{k=0}^{\tau-1} \pi_k$, the costs of observing until stopping time τ. For the payoff, we get

$$-\rho(\pi) = \sup \mathbb{E}[-(1 - \pi_\tau) - c \sum_{k=0}^{\tau-1} \pi_k] = -\inf \mathbb{E}[(1 - \pi_\tau) + c \sum_{k=0}^{\tau-1} \pi_k].$$

And therefore, $\rho(\pi) = \inf \mathbb{E}[(1 - \pi_\tau) + c \sum_{k=0}^{\tau-1} \pi_k]$ where the infimum is taking over the class of stopping times

$$\mathscr{M}^1 = \{\tau \in \mathscr{M} : \mathbb{E}[\sum_{k=0}^{\tau-1} \pi_k < \infty]\}.$$

It follows for the payoff

$$\begin{aligned} -\rho(\pi) &= \max\{g(x), T(-\rho(\pi)) - c(\pi_n)\} \\ &= \max\{-(1 - \pi), T(-\rho(\pi)) - \sum_{k=1}^{\tau-1} \pi_k)\} \qquad (4.23) \\ &= -\min\{(1 - \pi), c(\pi) + T\rho(\pi)\}, \end{aligned}$$

the payoff is the smallest excessive majorant of the function $g(\pi)$. $(Tf(x) - c(x) \geq f(x))$. We now get

$$\begin{aligned} \tau_0 &= \inf\{n \geq 0 : -\rho(\pi_n) \leq g(\pi)\}. \\ &= \inf\{n \geq 0 : -\min\{(1 - \pi), c(\pi) + T\rho(\pi)\} \leq -(1 - \pi_n)\} \\ &= \inf\{n \geq 0 : \min\{(1 - \pi), c(\pi) + T\rho(\pi)\} \geq 1 - \pi_n\} \end{aligned}$$

as the optimal stopping time to minimize our risk.

This is equivalent to the optimal stopping time $\tau_0 = \inf\{n \geq 0 : \rho(\pi_n) = 1 - \pi_n\}$

$$= \inf\{n \geq 0 : \pi_n \geq A^*\}.$$

Therefore, we found a optimal stopping time and proved Theorem 4.21.

□

4.5 Adding A Penalty For Choosing The Wrong Regime

In Shiryaev's book Shiryaev (2008) and later in Yakir's paper Yakir (1994) the risk function

$$\rho(\tau) = \mathbb{E}[I(\tau < \theta) + c_1 \mathbb{E}(\tau - \theta)^+)]$$

is used to find the change point. We showed that the risk function can be expressed with the posterior probability $\pi_\tau = P(\theta \leq \tau | \mathscr{F}_\tau)$ to make a decision while calculating the risk function. We previously found a change point detection method to a model with three regimes while we ignore the risk of choosing the wrong regime. The stopping rule is:

$$\tau_0 = \inf\{n \geq 0 : \pi_{n,B} + \pi_{n,C} \geq A^*\}. \tag{4.24}$$

Our goal is to develop a risk function that includes the penalty of choosing the wrong regime. A possible way to define that risk is:

$$\mathbb{E}[I(\theta \leq \tau)I(\pi_{\tau,B} \geq \pi_{\tau,C})I(\theta_C < \theta_B) + I(\theta \leq \tau)I(\pi_{\tau,C} > \pi_{\tau,B})I(\theta_C \geq \theta_B)]. \tag{4.25}$$

It describes the risk deciding on the wrong regime after a change point has happened. The decision criteria is based on the posterior functions $\pi_{\tau,B}$ and $\pi_{\tau,C}$. If $\pi_{\tau,B}$ is greater than or equal to $\pi_{\tau,C}$ we are deciding on regime B, otherwise regime C.

We can transform the risk into a formula depending on the posterior probabilities. We are showing this on the first term of our risk function.The risk will be updated by the information of our observations.

Lemma 4.26. *Let τ be a stopping time and $Y_\tau = I(\theta \leq \tau)I(\pi_{\tau,B} \geq \pi_{\tau,C})I(\theta_C < \theta_B) + I(\theta \leq \tau)I(\pi_{\tau,C} > \pi_{\tau,B})I(\theta_C \geq \theta_B)$ then*
$\mathbb{E}[Y_\tau] = E[I\{\pi_{\tau,B} \geq \pi_{\tau,C}\}\pi_{\tau,C} + I\{\pi_{\tau,C} > \pi_{\tau,B}\}\pi_{\tau,B}]$

Proof. First we are showing on the first term how to transform the risk into a formula depending on $\pi_{\tau,B}$.
$\mathbb{E}[I_{(\pi_{\tau,B} \geq \pi_{\tau,C})} I_{(\theta-1 \leq \tau, \theta_C < \theta_B)}] = \mathbb{E}[I(\pi_{\tau,B} > \pi_{\tau,C})\pi_{\tau,C}]$:

Therefore we get

$$\begin{aligned}
\mathbb{E}[I_{(\pi_{\tau,B} \geq \pi_{\tau,C})} I_{(\theta-1 \leq \tau, \theta_C < \theta_B)}] &= \mathbb{E}[I_{(\pi_{\tau,B} \geq \pi_{\tau,C})} I_{(\theta-1 \leq \tau, \alpha_\tau = C)}] \\
&= \mathbb{E}[\sum_n I_{(\pi_{n,B} \geq \pi_{n,C}, \theta-1 \leq n, \alpha_n = C)} I_{\tau=n}] \\
&= \sum_n \mathbb{E}[I_{(\pi_{n,B} \geq \pi_{n,C}, \theta-1 \leq n, \alpha_n = C)} I_{\tau=n}] \\
&= \sum_n \mathbb{E}[\mathbb{E}[I_{(\pi_{n,B} \geq \pi_{n,C})} I_{(\tau=n)} I_{(\theta-1 \leq n, \alpha_n = C)} | \mathscr{F}_n]] \qquad (4.27) \\
&= \sum_n \mathbb{E}[I_{(\pi_{n,B} \geq \pi_{n,C})} I_{(\tau=n)} \mathbb{E}[I_{(\theta-1 \leq n, \alpha_n = C)} | \mathscr{F}_n]] \\
&= \sum_n \mathbb{E}[I_{(\pi_{n,B} \geq \pi_{n,C})} \pi_{n,C} I_{(\tau=n)}] \\
&= \mathbb{E}[Y_\tau].
\end{aligned}$$

□

With Theorem 4.20, we need only to find the optimal stopping time for the problem:

$$\rho(\pi) = \inf \mathbb{E}[(1 - \pi_\tau) + c_1 \sum_{k=0}^{\tau-1} \pi_k + c_2(I\{\pi_{\tau,B} \geq \pi_{\tau,C}\}\pi_{\tau,C} + I\{\pi_{\tau,C} > \pi_{\tau,B}\}\pi_{\tau,B})]$$

and the infimum is taken over the stopping times.

$$\mathscr{M}^1 = \left\{ \tau \in \mathscr{M} : \mathbb{E}[\sum_{k=1}^{\tau-1} \pi_k] < \infty \right\}.$$

Therefore, we want to find the optimal stopping time while we are minimizing the risk:

$$R(\pi_{\tau,B}, \pi_{\tau,C}) = 1 - \pi_n + c_2 I\{\pi_{\tau,B} \geq \pi_{\tau,C}\}\pi_{\tau,C} + c_3 I\{\pi_{\tau,C} > \pi_{\tau,B}\}\pi_{\tau,B} + c_1 \sum_{k=1}^{\tau-1} \pi_k$$

We want minimize our risk function. Therefore we will be trying to maximize the negative risk:

$$-R(\pi_B, \pi_C) = -(1 - \pi_n) - c_2(I\{\pi_{\tau,B} \geq \pi_{\tau,C}\}\pi_{\tau,C} - I\{\pi_{\tau,C} > \pi_{\tau,B}\}\pi_{\tau,B}) - c_1 \sum_{k=1}^{\tau-1} \pi_k.$$

We have

$$g(\pi_B, \pi_C) = -(1 - (\pi_B + \pi_C)) - c_2 I_{(\pi_B \geq \pi_C)}\pi_C - c_3 I_{(\pi_B < \pi_C)}\pi_B$$

and

$$c(\pi_B, \pi_C) = \pi.$$

It can be shown that $|g(\pi_B, \pi_C)| < G < \infty$ and $\mathbb{E}[c(\pi_{n,B}, \pi_{n,C})] < \infty$.

Shiryaev defines the pay of function as

$$s(x) = \sup \mathbb{E} \left\{ \alpha^\tau g(x_\tau) - \sum_{s=0}^{\tau-1} \alpha^s c(x_s) \right\}.$$

Comparing this to our negative risk function we get:

$$s(\pi_B,\pi_C)=\sup\mathbb{E}\left\{g(\pi_{\tau,B},\pi_{\tau,C})-\sum_{s=0}^{\tau-1}c(\pi_s)\right\}$$
$$=\sup\mathbb{E}\left\{-(1-\pi_\tau)-c_2I\{\pi_{\tau,B}\geq\pi_{\tau,C}\}\pi_{\tau,C}-c_3I\{\pi_{\tau,C}>\pi_{\tau,B}\}\pi_{\tau,B}-c_1\sum_{s=0}^{\tau-1}\pi_s\right\}$$
$$=-\inf\mathbb{E}\left\{(1-\pi_\tau)+c_2I\{\pi_{\tau,B}\geq\pi_{\tau,C}\}\pi_{\tau,C}+c_3I\{\pi_{\tau,C}>\pi_{\tau,B}\}\pi_{\tau,B}+c_1\sum_{s=0}^{\tau-1}\pi_s\right\}.$$

The last equation shows that $s(\pi_B,\pi_C)=-\rho(\pi_B,\pi_C)$.

Therefore according to Theorem 4.20 $s(x)$ is the smallest (α,c)-excessive majorant of the function $g(x)$.

$$s(\pi_B,\pi_C)=\max\{g(\pi_B,\pi_C),Ts(\pi_B,\pi_C)-c(\pi_B,\pi_C)\}$$
$$=\max\left\{-(1-\pi)-c_2I_{(\pi_B\geq\pi_C)}\pi_C-c_3I_{(\pi_B<\pi_C)}\pi_B,Ts(\pi_B,\pi_C)-\pi\right\}$$

and

$$\rho(\pi_B,\pi_C)=-\max\left\{-(1-\pi)-c_2I_{(\pi_B\geq\pi_C)}\pi_C-c_3I_{(\pi_B<\pi_C)}\pi_B,Ts(\pi_B,\pi_C)-\pi\right\}$$
$$=\min\left\{(1-\pi)+c_2I_{(\pi_B\geq\pi_C)}\pi_C+c_3I_{(\pi_B<\pi_C)}\pi_B,\pi-Ts(\pi_B,\pi_C)\right\}$$
$$=\min\left\{(1-\pi)+c_2I_{(\pi_B\geq\pi_C)}\pi_C+c_3I_{(\pi_B<\pi_C)}\pi_B,\pi+T\rho(\pi_B,\pi_C)\right\}$$

where $Ts(\pi_{n,B},\pi_{n,C})=\mathbb{E}[s(\pi_{B,1},\pi_{C,1})]$.

Furthermore, since $P(\tau_0<\infty)=1$ and $P(\sum_{s=0}^{\infty}c(\pi)=\infty)=1$, the optimal stopping time is equal to

$$
\begin{aligned}
\tau_0 &= \inf\{n \geq 0 : s(\pi_{n,B}, \pi_{n,C}) \leq g(\pi_{n,B}, \pi_{n,C}))\} \\
&= \inf\{n \geq 0 : \max\left\{-(1-\pi_n) - c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} - I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}, Ts(\pi_B, \pi_C) - \pi_n\right\} \\
&\leq -(1-(\pi_{n,B}+\pi_{n,C})) - c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} - c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\} \\
&= \inf\{n \geq 0 : -\min\left\{(1-\pi_n) + c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}, \pi - Ts(\pi_{n,B}, \pi_{n,C})\right\} \\
&\leq -(1-(\pi_{n,B}+\pi_{n,C})) - c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} - c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\} \\
&= \inf\{n \geq 0 : \min\left\{(1-\pi_n) + c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}, \pi - Ts(\pi_{n,B}, \pi_{n,C})\right\} \\
&\geq (1-(\pi_{n,B}+\pi_{n,C})) + I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\} \\
&= \inf\{n \geq 0 : \min\left\{(1-\pi_n) + c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}, \pi + T\rho(\pi_{n,B}\pi_{n,C})\right\} \\
&\geq (1-(\pi_{n,B}+\pi_{n,C})) + c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\} \\
&= \inf\{n \geq 0 : \rho(\pi_{n,B}, \pi_{n,C}) \geq (1-(\pi_{n,B}+\pi_{n,C})) + c_2 I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\}.
\end{aligned}
$$

As a result our optimal stopping time is

Theorem 4.28. *The optimal stopping time for the model described in Section 4.1 with a penalty for choosing the wrong regime is*

$$
\tau_0 = \inf\{n \geq 0 : \rho(\pi_{n,B}, \pi_{n,C}) \geq (1-(\pi_{n,B}+\pi_{n,C})) + I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\}.
$$

5.

Estimating The Risk ρ

In this chapter we are discussing the first steps of the calculations of our change point detection method. First, we are showing on a simple example that the calculations for ρ are computationally intensive (Section 5.1). Therefore, an estimation for the risk ρ is derived in the following sections. Here, we are showing how this estimation can be applied on different models.

5.1 Change Point Detection Example

We are looking at a Markov chain with two possible states, 0 and 1 where our current state is denoted as x_i with an initial state $x_0 = 0$. In addition, we have three transition matrices A,B and C.

$$A = \begin{pmatrix} 0.1 & 0.9 \\ 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 0 \\ 0.9 & 0.1 \end{pmatrix}$$

$$C = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

Our regime A has a transition probability matrix that makes transitions to state 1 very likely. In contrast, regime B tends to have more transitions towards stage 0. The third regime has even probabilities.

Our change points are geometrically distributed with $\theta_B \sim \text{geom}(0.3)$ and $\theta_C \sim \text{geom}(0.3)$. In addition, the probabilities that a change point will happen before our first observation, are $\pi_B = 0$ and $\pi_C = 0$.

With the results from Lemma 4.4, the posterior probabilities can be calculated as follows:

$$\pi_{n,B} = \frac{\pi_{n-1,B}\ b_{n-1,n} + (1-(\pi_{n-1,B}+\pi_{n-1,C}))\ \lambda_B\ a_{n-1,n}}{numerator_B + numerator_C +\ (1-(\pi_{n-1,B}+\pi_{n-1,C}))\ q\ a_{n-1,n}}$$

$$\pi_{n+1,B} = \frac{\pi_{n,B} b_{n-1,n} + (1-(\pi_{n,B}+\pi_{n,C}))\ 0.3\ a_{n,n+1}}{\text{numerator}_B + \text{numerator}_C + (1-(\pi_{n,B}+\pi_{n,C}))\ 0.49\ a_{n,n+1}}$$

with $q = 1-p = 1-(\lambda_B+\lambda_C-(\lambda_B\lambda_C)) = 1-(0.3+0.3-(0.3\times 0.3)) = 0.49$.

For regime C:

$$\pi_{n+1,C} = \frac{\pi_{n,C} c_{n-1,n} + (1-(\pi_{n,B}+\pi_{n,C}))\ 0.21\ a_{n,n+1}}{\text{numerator}_B + \text{numerator}_C + (1-(\pi_{n,B}+\pi_{n,C}))\ 0.49\ a_{n,n+1}}$$

with $\lambda_C(1-\lambda_B) = 0.21$.

After running the simulations with the given parameters, we are getting the values $x_0, \ldots, x_9$ shown in Table 5.1.

The calculated posterior values are shown in Table 5.2. The first array represents the

i	0	1	2	3	4	5	6	7	8	9
x_i	0	1	1	0	1	0	1	1	1	0

Table 5.1: First ten values for a simulation of the Markov Chain with two alternative regimes with a switch to regime C at time step 2

i	0	1	2	3	4	5	6	7	8	9
$\pi_{n,B}$	0	0.3	0.2832	0.2480	0.5113	0.0	0.0	0.0	0.0	0.0
$\pi_{n,C}$	0.0	0.21	0.3326	0.4267	0. 4887	1.0	1.0	1.0	1.0	1.0

Table 5.2: Calculated Posterior Values for n simulations

posterior probabilities for regime B at time steps 0 to 9 and the second array shows the posterior probabilities of regime C.

We want to make a decision based on our optimal stopping time. Therefore, we declare that a change point has happened if

$$\min\left\{(1-\pi_n) + c_2 I_{(\pi_{n,B}\geq\pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B}<\pi_{n,C})}\pi_{n,B},\ \pi + T\rho(\pi_{n,B}\pi_{n,C})\right\}$$
$$\geq (1-\pi_n) + c_2 I_{(\pi_{n,B}\geq\pi_{n,C})}\pi_{n,C} + c_3 I_{(\pi_{n,B}<\pi_{n,C})}\pi_{n,B}$$

which is analog to the method developed in Section 4.4. Using the calculated posterior functions, the first step in calculations gives us:

$$\min\left\{(1-\pi_1) + c_2 I_{(\pi_{B,1}\geq\pi_{C,1})}\pi_{C,1} + c_3 I_{(\pi_{B,1}<\pi_{C,1})}\pi_{B,1},\ \pi + T\rho(\pi_{B,2}\pi_{C,2})\right\}$$
$$\geq (1-(\pi_{B,1}+\pi_{C,1})) + c_2 I_{(\pi_{B,1}\geq\pi_{C,1})}\pi_{C,1} + c_3 I_{(\pi_{B,1}<\pi_{C,1})}\pi_{B,1}$$
$$\geq \min\{0.49+0.21, 0.51+\mathbb{E}[\pi_{2,B},\pi_{2,C}]\} \geq 0.49+0.21$$
$$\geq \min\{0.7, 0.51+\mathbb{E}[\pi_{2,B},\pi_{2,C}]\} \geq 0.7.$$

Looking at the stopping rule at time step 1, we have to do more calculations. The declaration of a change point depends on the calculations of $\mathbb{E}[\rho(\pi_{2,B}, \pi_{2,C})]$. Therefore, we have to calculate the expected values until we know that the first term 0.7 is the minimum. In order to calculate the expected risk for making another observation, we are calculating the possible values for $\rho(\pi_{2,B}, \pi_{2,C})$ and their probabilities.

If $x_2 = 0$, we are getting the posterior probabilities $\pi_{2,B} = 0.72$ and $\pi_{2,C} = 0.28$.

For the case $x_2 = 1$, we get $\pi_{2,B} = 0.2832$ and $\pi_{2,C} = 0.3326$.

In our case, the probability that $P(x_2|x_1, x_0)$ can be calculated as follows:

$$\begin{aligned}
&P(x_2|x_1, x_0) = P(x_2|x_1, x_0, \alpha_1 = A)\ P(\alpha_1 = A|x_1, x_0) \\
&\quad + P(x_2|x_1, \alpha_1 = B)\ P(\alpha_1 = B|x_1, x_0) + P(x_2|x_1, \alpha_1 = C)\ P(\alpha_1 = C|x_1, x_0) \\
&= a(x_2|x_1)P(\alpha_1 = A|x_1, x_0) + b(x_2|x_1)P(\alpha_1 = B|x_1, x_0) + c(x_2|x_1)P(\alpha_1 = C|x_1, x_0) \\
&= a(x_2|x_1)(1 - (\pi_{1,B} + \pi_{1,C}) + b(x_2|x_1)\pi_{1,B} + c(x_2|x_1)\pi_{1,C} \\
&= a(x_2|x_1)(1 - (\pi_{1,B} + \pi_{1,C}) + b(x_2|x_1)\pi_{1,B} + c(x_2|x_1)\pi_{1,C}.
\end{aligned}$$

The calculated probabilities for x_2 are

$$\begin{aligned}
&P(x_2 = 0|x_1 = 1, x_0 = 0) = 0 \times (1 - (0.3 + 0.21)) + 0.9 \times 0.3 + 0.5 \times 0.21 \\
&= 0.375
\end{aligned}$$

and

$$\begin{aligned}
&P(x_2 = 1|x_1 = 1, x_0 = 0) = 1 \times (1 - (0.3 + 0.21)) + 0.1 \times 0.3 + 0.5 \times 0.21 \\
&= 0.625.
\end{aligned}$$

We can now calculate $\mathbb{E}[\rho(\pi_{2,B}, \pi_{2,C})]$:

$$\mathbb{E}[\rho(\pi_{2,B}, \pi_{2,C})] = \sum_{\rho(\pi_{2,B}, \pi_{2,C})} \rho(\pi_{2,B}, \pi_{2,C}) P(\rho(\pi_{2,B}, \pi_{2,C})|x_1, x_0)$$

$$= \sum_{x_2} \rho(\pi_{2,B}, \pi_{2,C}) P(x_2|x_1, x_0)$$

$$= \rho(\pi_{2,B}, \pi_{2,C}) P(x_2 = 0|x_1, x_0) + \rho(\pi_{2,B}, \pi_{2,C}) P(x_2 = 1|x_1, x_0)$$

$$= \rho(\pi_{2,B}, \pi_{2,C}) P(x_2 = 0|x_1 = 0, x_0 = 1) + \rho(\pi_{2,B}, \pi_{2,C}) P(x_2 = 1|x_1 = 0, x_0 = 1)$$

$$= \rho(\pi_{2,B}, \pi_{2,C}) 0.375 + \rho(\pi_{2,B}, \pi_{2,C}) 0.625.$$

The risk function $\rho(\pi_{2,B}, \pi_{2,C})$ can be calculated for both cases $x_2 = 0$ and $x_1 = 1$.

For $x_2 = 0$:

$$\rho(\pi_{2,B}, \pi_{2,C}) = \min\{1 - (0.72 + 0.28) + 0.28, 0.72 + 0.28 + T\rho(\pi_{3,B}, \pi_{3,B})\}$$

$$= \min\{0.28, 1 + T\rho(\pi_{3,B}, \pi_{3,B})\}$$

since $T\rho(\pi_{3,B}, \pi_{3,B})$ is positive, $\rho(\pi_{2,B}, \pi_{2,C}) = 0.28$.

For $x_2 = 1$:

$$\rho(\pi_{2,B}, \pi_{2,C}) = \min\{1 - (0.2832 + 0.3326) + 0.2832, 0.2832 + 0.3326 + T\rho(\pi_{3,B}, \pi_{3,B})\}$$

$$= \min\{0.6674, 0.6158 + T\rho(\pi_{3,B}, \pi_{3,B})\}.$$

We cannot state the exact value of $\rho(\pi_{2,B}, \pi_{2,C})$ because we do not know what $T\rho(\pi_{3,B}, \pi_{3,B})$

i	x_i	α_i	$\pi_{i,B}$	$\pi_{i,C}$	ρ_{left}	ρ_{right}	ρ	change point
1	1	A	0.3	0.21	0.7	0.5221	0.5221	
2	1	C	0.2832	0.3326	0.6674	0.4981	0.4981	
3	0	C	0.2480	0.4267	0.5733	0.4951	0.4951	
4	1	C	0.5113	0.4887	0.4887	0.2443	0.2443	
5	0	C	0.0	1.0	0		0	$\tau_0 = 5$
6	1	C	0.0	1.0	0		0	
7	1	C	0.0	1.0	0		0	
8	1	C	0.0	1.0	0		0	
9	0	C	0.0	1.0	0		0	

Table 5.3: Values for simulated Markov chain

is. With these calculations we get:

$$\mathbb{E}[\rho(\pi_{2,B}, \pi_{2,C})] = \rho(\pi_{2,B}, \pi_{2,C})0.375 + \rho(\pi_{2,B}, \pi_{2,C})0.625$$

$$= 0.28 \times 0.375 + \min\{0.6674, 0.6158 + T\rho(\pi_{3,B}, \pi_{3,B}) \times 0.625.$$

Our stopping rule for $x_1 = 1$ is:

$$\begin{aligned}
\rho(\pi_{1,B}, \pi_{1,C}) &= \min\{0.7, 0.51 + \mathbb{E}[\pi_{2,B}, \pi_{2,C}]\} \\
&= \min\{0.7, 0.51 + 0.28 \times 0.375 \\
&\quad + \min\{0.6674, 0.6158 + T\rho(\pi_{3,B}, \pi_{3,B})\} \times 0.625\} \\
&= \min\{0.7, 0.51 + 0.105 + \min\{0.6674, 0.6158 + 0.4981\} \times 0.625\} \\
&= \min\{0.7, 0.51 + 0.105 + 0.6674 \times 0.625\} \\
&= \min\{0.7, 0.5221\} \\
&= 0.5221.
\end{aligned}$$

Therefore, we are not causing an alarm. The following calculations can be found in Table 5.3. A change point is detected after a switch from regime 1 to 0. This is only possible under a regime change to C. The alarm is caused at time step $\tau_0 = 5$.

5.2 Using Q^n To Estimate ρ

The results of Section 5.1 show that calculating ρ is computationally intensive. Therefore, we are interested to use a risk Q^n as in Theorem 4.20, to estimate the risk ρ. In this section we derive the risk Q for a Markov chain model with two regimes and show its application on a simulation.

5.2.1 Calculating Q for Markov Chain Model With Two Regimes

After calculating Q^n for a simple model with one alternative regime, we are now looking at the case for Markov chain random variables. We assume that $X_1, X_2, \dots$ are random variables governed by a Markov chain distribution with

$$P(X_0 = x) = 1$$

$$P(x_i|x_{i-1}) = a(x_i|x_{i-1})$$

for $i = 1, 2, \dots, \theta - 1$, and

$$P(x_i|x_{i-1}) = b(x_i|x_{i-1})$$

for $i = \theta, \theta + 1, \dots$. Here θ is the change point with zero inflated distribution

$$P(\theta = 0) = \pi \in (0, 1)$$

$$P(\theta = n) = (1 - \pi)(1 - p)^{n-1}p, \quad n = 1, 2, \dots$$

where $p \in (0,1)$.

The posterior probability is defined as

$$\pi_n := P(\theta \leq n | \mathscr{F}_n), n = 0, 1, 2, \dots$$

where

$$\mathscr{F}_0 = \{\emptyset, \Omega\}$$

$$\mathscr{F}_n = \sigma\{X_1, X_2, \dots, X_n\}.$$

Some observations that are

$$\pi_0 = P(\theta \leq 0 | \mathscr{F}_0) = P(\theta = 0) = \pi$$

$$\pi_1 = P(\theta \leq 1 | \mathscr{F}_1) = P(\theta = 0 | \mathscr{F}_1) + P(\theta = 1 | \mathscr{F}_1).$$

Then

$$\begin{aligned} \mathbb{E}[\pi_1] &= \mathbb{E}[P(\theta = 0 | \mathscr{F}_1) + P(\theta = 1 | \mathscr{F}_1)] \\ &= P(\theta = 0) + P(\theta = 1) \\ &= \pi + (1 - \pi) \cdot p \\ &= p + (1 - p) \cdot \pi. \end{aligned} \tag{5.1}$$

We derive a Bayes formula for π_1 following Shiryaev's approach Shiryaev (2008). Consider $A = \{X_1 = x_1\} \in \mathscr{F}_1$. Then

$$\begin{aligned} P_x(\theta \leq 1, A) &= P(\theta = 0, A | X_0 = x) + P(\theta = 1, A | X_0 = x) \\ &= P(A | \theta = 0, X_0 = x) \cdot P(\theta = 0) + P(A | \theta = 1, X_0 = x) \cdot P(\theta = 1) \\ &= \pi P(X_1 = x_1 | \theta = 0, X_0 = x) + (1 - \pi) p P(X_1 = x_1 | \theta = 1, X_0 = x) \\ &= \pi \cdot b(x_1 | x) + (1 - \pi) \cdot p \cdot b(x_1 | x). \end{aligned} \tag{5.2}$$

We use the definition of conditional expectation to derive a Bayes formula for π_1 and define

$$\begin{aligned}P_x(X_1 = x_1) &= P(A, \theta = 0) + P(A, \theta = 1) + P(A, \theta > 1) \\ &= \pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi) \cdot (1-p) \cdot a(x_1|x).\end{aligned} \tag{5.3}$$

Therefore,

$$\begin{aligned}P_x(\theta \leq 1, A) &= \pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) \\ &= \frac{\pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x)}{\pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi) \cdot (1-p) \cdot a(x_1|x)} \cdot P(A) \\ &= \int_A \frac{\pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x1|x)}{\pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi) \cdot (1-p) \cdot a(X_1|x)} dP \\ &= \int_A \frac{b(x_1|x) \cdot (\pi + (1-\pi) \cdot p)}{\pi \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi) \cdot (1-p) \cdot a(x_1|x)} dP.\end{aligned} \tag{5.4}$$

By the definition of conditional expectation, we have

$$P_x(\theta \leq 1|\mathscr{F}_1) = \pi_1(X_1) = \frac{\pi \cdot b(X_1|x) + (1-\pi) \cdot p \cdot b(X_1|x)}{\pi \cdot b(X_1|x) + (1-\pi) \cdot p \cdot b(X_1|x) + (1-\pi) \cdot (1-p) \cdot a(X_1|x)} \tag{5.5}$$

given that $\pi_1(X_1)$ is $\mathscr{F}_1$-measurable.

Moreover,

$$\begin{aligned}\mathbb{E}_x[\pi_1] &= \sum \pi_1(x_1) P(X_1 = x_1|X_0 = x) \\ &= \sum \frac{\pi_0 \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x)}{\pi_0 \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi)(1-p) a(x_1|x)} \\ &\cdot (\pi_0 \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) + (1-\pi)(1-p) a(x_1|x)) \\ &= \sum \pi_0 \cdot b(x_1|x) + (1-\pi) \cdot p \cdot b(x_1|x) \\ &= \pi_0 + (1-\pi_0) \cdot p.\end{aligned} \tag{5.6}$$

We can see that $\mathbb{E}_x[\pi_1]$ does not depend on x which agrees with the previous results where we calculated the expected value directly.

5.2.2 Example: Calculating Q^1 for Markov Chain Model With Two Regimes

We assume $x_1, x_2 \dots$ are random variables governed by a Markov Chain distribution with $P(x_0 = 0) = 1$ The transition probability matrices are

$$A = \begin{pmatrix} 0.1 & 0.9 \\ 0.4 & 0.6 \end{pmatrix}$$

for $i = 1, 2, \dots, \theta - 1$ and

$$B = \begin{pmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{pmatrix}$$

for $i = \theta, \theta + 1, \dots$.

and θ is the change point with distribution

$$P(\theta = 0) = \pi \in (0, 1)$$

$$P(\theta = n) = (1 - \pi) \cdot (1 - p)^{n-1} \cdot p.$$

Let $p = 0.4$

We are calculating $Q(g(\pi)) = \min\{g(\pi), c \cdot \pi + \mathbb{E}g(\pi)\}$ for $\pi = 0.1$ and $c = 1$. We have $g(\pi) = 1 - \pi = 0.9$ and $c \cdot \pi = 0.1$.

Furthermore,

$$\begin{aligned} P(x_1 = 1 | x_0 = 0) &= \pi \cdot b(1|0) + (1 - \pi)(p)b(1|0) + (1 - \pi)(1 - p)a(1|0) \\ &= 0.1 \cdot 0.2 + 0.9 \cdot 0.4 \cdot 0.2 + 0.9 \cdot 0.6 \cdot 0.9 \qquad (5.7) \\ &= 0.578 \end{aligned}$$

and

$$P(x_1 = 0|x_0 = 0) = \pi \cdot b(0|0) + (1-\pi)(p)b(0|0) + (1-\pi)(1-p)a(0|0)$$
$$= 0.1 \cdot 0.8 + 0.9 \cdot 0.4 \cdot 0.8 + 0.9 \cdot 0.6 \cdot 0.1 \tag{5.8}$$
$$= 0.422.$$

In addition, we can calculate $\mathbb{E}[\pi_1] = \mathbb{E}[\pi_1|\pi_1 = 0]P(x_1 = 0|P_0 = 0) + \mathbb{E}[\pi_1|\pi_1 = 1]P(x_1 = 1|P_0 = 0)$:

$$\begin{aligned} \mathbb{E}[\pi|x_1 = 0] &= \frac{\pi \cdot b(0|0) + (1-\pi)(p)b(0|0)}{\pi \cdot b(0|0) + (1-\pi)(p)b(0|0) + (1-\pi)(1-p)a(0|0)} \\ &= \frac{0.1 \cdot 0.2 + 0.9 \cdot 0.4 \cdot 0.2}{0.1 \cdot 0.2 + 0.9 \cdot 0.4 \cdot 0.2 + 0.9 \cdot 0.6 \cdot 0.9} \\ &= \frac{0.092}{0.578} \\ &= 0.1592 \end{aligned} \tag{5.9}$$

$$\begin{aligned} \mathbb{E}[\pi|x_1 = 1] &= \frac{\pi \cdot b(1|0) + (1-\pi)(p)b(1|0)}{\pi \cdot b(1|0) + (1-\pi)(p)b(0|0) + (1-\pi)(1-p)a(0|0)} \\ &= \frac{0.1 \cdot 0.8 + 0.9 \cdot 0.4 \cdot 0.8}{0.1 \cdot 0.8 + 0.9 \cdot 0.4 \cdot 0.8 + 0.9 \cdot 0.6 \cdot 0.1} \\ &= \frac{0.368}{0.422} \\ &= 0.8720 \end{aligned} \tag{5.10}$$

We get $\mathbb{E}[g(\pi)] = 0.578 \cdot (1 - 0.1592) + 0.422 \cdot (1 - 0.8720) = 0.540$ and

$$Q(g(\pi)) = \min\{0.9, 0.1 + 0.540\} = 0.640$$

.

Similarly, we calculated values for different π values as shown in Table 5.4.

π	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$Qg(\pi)$	0.6	0.64	0.68	0.7	0.6	0. 5	0.4	0.3	0.2	0.1	0.0

Table 5.4: $Qg(\pi)$ values for a Markov Chain model

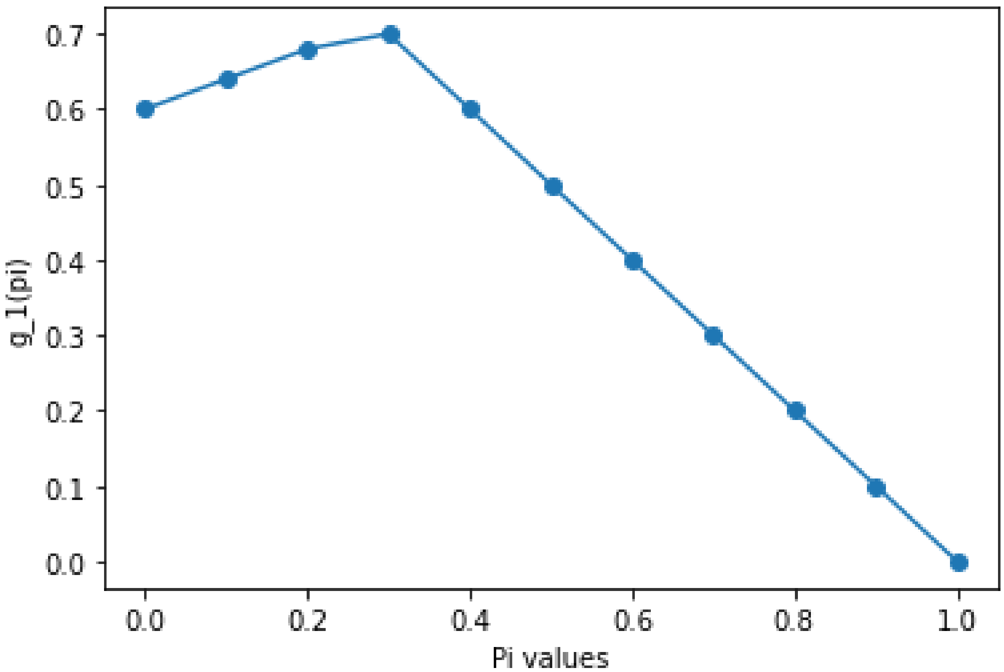


Figure 5.1: Interpolation between the calculated values of $Qg(\pi)$

5.2.3 Example: Calculating $Q^2(\pi)$ for Markov Chain Model With Two Regimes

After calculating $g_1 = Q(g(\pi))$ we are interested in calculating $g_2(\pi) = Q^2$.

$$\begin{aligned} g_2(\pi) &= \min\{g_1(\pi), \pi + \mathbb{E}[Q(\pi)]\} \\ &= \min\{0.640, 0.1 + \mathbb{E}[Q(\pi)]\} \end{aligned} \tag{5.11}$$

All values are given besides $\mathbb{E}[Q(\pi)]$. Therefore, we can use the $Q(\pi)$ value that we calculated previously by interpolating between the values. The function that is achieve by interpolating between the points is shown in (5.1).

π	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$Qg(\pi)$	0.6	0.64	0.68	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
$Q^2(\pi)$	0.464	0.5376	0.6112	0.6814	0.6	0.5	0.4	0.3	0.2	0.1	0.0

Table 5.5: $Qg(\pi)$ values for a Markov Chain model

$$\begin{aligned}\mathbb{E}_\pi[g_1(\pi)] &= g_1(\pi) \cdot P_\pi(x_1 = 0|x_0 = 0) + g_1(\pi) \cdot P_\pi(x_1 = 1|x_0 = 0) \\ &= g_1(0.1592) \cdot 0.578 + g_1(0.8720) \cdot 0.422 \\ &= 0.6638 \cdot 0.578 + 0.128 \cdot 0.422 \\ &= 0.4377\end{aligned} \tag{5.12}$$

Therefore, we get the following value for $g_2(0.1)$:

$$\begin{aligned}g_2(\pi) &= \min\{0.640, 0.1 + 0.4377\} \\ &= 0.5377\end{aligned} \tag{5.13}$$

Further calculations with other π values give us an overview of the function $Q^2(\pi)$ shown in Table 5.5.

A comparison of the different Q^n and its convergence to a function that will later be called $\rho(\pi)$ is shown in graphic Figure 5.2. The blue line shows our first $Q(\pi)$, the orange line represents Q^2. The other two lines are not discernible on the graph. They show Q^{100} and Q^{1000}. It can be seen that the values do not change significantly anymore. Therefore, we can call the red line and estimate of $\rho(\pi)$.

5.3 Markov Chain Model With Three Regimes

In order to use the derived stopping time method, we have to calculate ρ. However, we showed that the calculations become computational intensive fast even for simpler models. Therefore, we derive a method to estimate ρ for a Markov Chain with three regimes and show its application on an example. In the following chapter we are adapting the parameter Q derived previously.

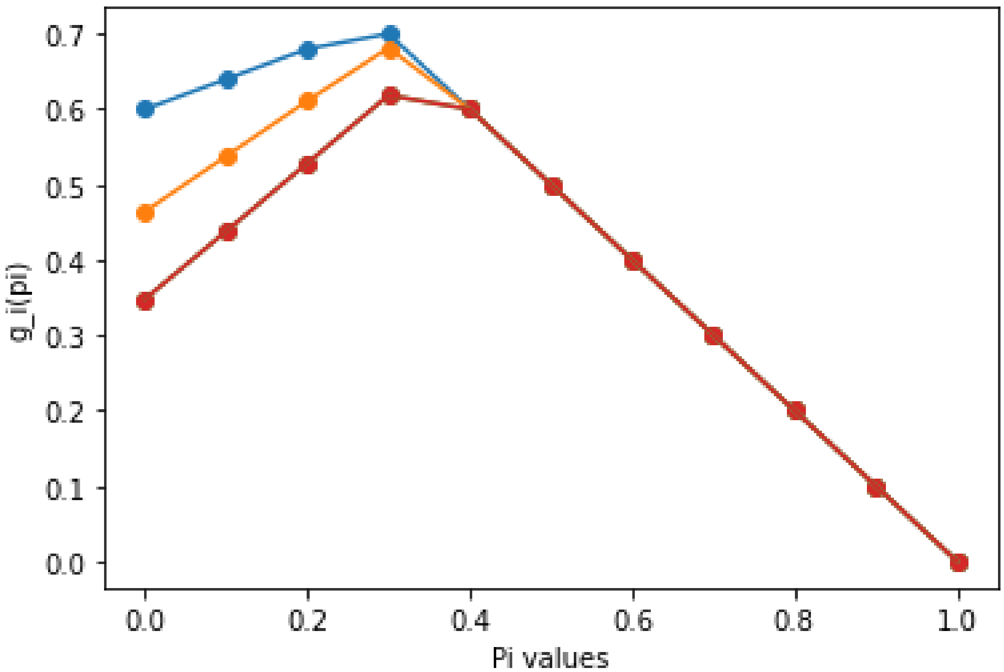


Figure 5.2: Interpolation between the calculated values of $Qg(\pi)$ with blue: $Q(\pi)$, orange: Q^2 and red: Q^{1000}.

5.3.1 Q^n for a Markov Chain Model With Three Regimes

Our goal is to derive a method that detects a change point with two alternative regimes. We again assume $X_1, X_2, \ldots$ are random variables governed by a Markov chain distribution with

$$P(X_0 = x) = 1$$

$$P(x_i|x_{i-1}) = a(x_i|x_{i-1})$$

for $i = 1, 2, \ldots, \theta - 1$. After the change point $\theta = \min\{\theta_B, \theta_C\}$ where $\theta_B \sim \text{geom}(\lambda_B)$ and $\theta_C \sim \text{geom}(\lambda_C)$ we have

$$P(x_i|x_{i-1}) = b(x_i|x_{i-1})$$

if $\theta = \theta_B$ and

$$P(x_i|x_{i-1}) = c(x_i|x_{i-1})$$

if $\theta = \theta_C$

for $i = \theta, \theta + 1, \ldots$

Therefore we have:

$$P(\theta = 0) = P(\theta_B = 0) + P(\theta_C = 0 \cap \theta_B > 0) = \pi_B + (1 - \pi_B) \cdot \pi_C = \pi_0 \ \in (0, 1)$$

and

$$P(\theta = n) = (1 - \pi_0)(1 - p)^{n-1} p, n = 1, 2, \ldots$$

where $p = \lambda_B + \lambda_C - \lambda_B \cdot \lambda_C$.

The posterior probabilities are defines as

$$\pi_{n,B} := P(\theta \leq n, \alpha_n = B | \mathscr{F}_n)$$

and

$$\pi_{n,C} := P(\theta \leq n, \alpha_n = C | \mathscr{F}_n).$$

The first observations are

$$\begin{aligned} \pi_0 &= P(\theta \leq 0 | \mathscr{F}_0) \\ &= P(\theta \leq 0, \alpha_0 = B | \mathscr{F}_0) + P(\theta \leq 0, \alpha_0 = C | \mathscr{F}_0) \\ &= P(\theta_B = 0) + (1 - P(\theta_B = 0)) \cdot P(\theta_C = 0) \\ &= \pi_B + (1 - \pi_B) \cdot \pi_C \end{aligned} \tag{5.14}$$

$$\pi_1 = P(\theta \leq 1 | \mathscr{F}_1) = P(\theta = 0 | \mathscr{F}_1) + P(\theta = 1 | \mathscr{F}_1). \tag{5.15}$$

Then

$$
\begin{aligned}
\mathbb{E}[\pi_1] &= \mathbb{E}[P(\theta = 0|\mathscr{F}_1) + P(\theta = 1|\mathscr{F}_1)] \\
&= P(\theta = 0) + P(\theta = 1) \\
&= \pi_0 + (1 - \pi_0) \cdot p \\
&= p + (1 - p) \cdot \pi_0.
\end{aligned} \tag{5.16}
$$

We derive a Bayes formula for π_1, as we did previously. Consider $A = \{X_1 = x_1\} \in \mathscr{F}_1$. Then

$$
\begin{aligned}
P_x(\theta \leq 1, A) &= P(\theta = 0, A|X_0 = x) + P(\theta = 1, A|X_0 = x) \\
&= P(\theta = 0, \alpha_1 = B, A|X_0 = x) + P(\theta = 0, \alpha_1 = C, A|X_0 = x) \\
&\quad + P(\theta = 1, \alpha_1 = B, A|X_0 = x) + P(\theta = 1, \alpha_1 = C, A|X_0 = x) \\
&= P(A|\theta = 0, \alpha_1 = B, X_0 = x) \cdot \pi_B \\
&+ P(A|\theta = 0, \alpha_1 = C, X_0 = x) \cdot (1 - \pi_B) \cdot \pi_C \\
&\quad + P(A|\theta = 1, \alpha_1 = B, X_0 = x) \cdot P(\theta = 1, \alpha_1 = B) \\
&\quad + P(A|\theta = 1, \alpha_1 = C, X_0 = x) \cdot P(\theta = 1, \alpha_1 = C) \\
&= \pi_B \cdot b(x_1|x) + (1 - \pi_B) \cdot \pi_C \cdot c(x_1|x) \\
&\quad + (1 - \pi_0) \cdot \lambda_B \cdot b(x_1|x) \\
&\quad + (1 - \pi_0) \cdot (1 - \lambda_B) \cdot \lambda_C \cdot c(x_1|x).
\end{aligned} \tag{5.17}
$$

We apply the definition for the conditional expectation $\int_G X dP = \int_G \mathbb{E}[X|\mathscr{G}] dP$ in order to derive the Bayes formula.

$$\begin{aligned}P_x(A) &= P_x(X_1 = x_1)\\&= \pi_B \cdot b(x_1|x) + (1-\pi_B)\cdot \pi_C \cdot c(x_1|x) + (1-\pi_0)\cdot \lambda_B \cdot b(x_1|x)\\&+ (1-\pi_0)\cdot(1-\lambda_B)\cdot \lambda_C \cdot c(x_1|x)\\&+ (1-\pi_0)\cdot(1-\lambda_B)\cdot(1-\lambda_C)\cdot a(x_1|x)\end{aligned} \tag{5.18}$$

$$\begin{aligned}P_x(\theta \le 1, A) &= \frac{\pi_B b(x_1|x) + (1-\pi_B)\pi_C c(x_1|x)}{P(A)}\\&+ \frac{(1-\pi_0)\left[\lambda_B b(X_1|x) + (1-\lambda_B)\lambda_C c(x_1|x)\right]}{P(A)} \cdot P(A)\\&= \int_A \frac{\pi_B b(x_1|x) + \pi_C c(x_1|x) + (1-\pi_0)\left[\lambda_B b(X_1|x) + (1-\lambda_B)\lambda_C c(x_1|x)\right]}{P(A)} dP.\end{aligned} \tag{5.19}$$

We can use the definition of the conditional expectation, given that π_1 is $\mathscr{F}_1$-measurable.

$$\int_A \frac{\pi_B b(x_1|x) + \pi_C c(x_1|x) + (1-\pi_0)\left[\lambda_B b(X_1|x) + (1-\lambda_B)\lambda_C c(x_1|x)\right]}{P(A)} dP = \int_A \pi_1 dP. \tag{5.20}$$

This result gives us a Bayes formula for π_1:

$$\pi_{x,1} = \frac{\pi_B b(X_1|x) + \pi_C c(X_1|x) + (1-\pi_0)\left[\lambda_B b(X_1|x) + (1-\lambda_B)\lambda_C c(X_1|x)\right]}{P(X_1)}. \tag{5.21}$$

We can use this formula to calculate $\mathbb{E}[\pi_1]$ which is necessary to calculate different values of

Q.

$$\begin{aligned}
\mathbb{E}_x[\pi_1] &= \sum \pi_1(x_1) \cdot P(X_1 = x_1 | X_0 = x) \\
&= \sum \frac{\pi_B \cdot b(x_1|x) + (1-\pi_B) \cdot \pi_C \cdot c(x_1|x)}{P(X_1 = 1)} \\
&+ \frac{(1-\pi_0) \cdot [\lambda_B b(x_1|x) + (1-\lambda_B) \cdot \lambda_C \cdot c(x_1|x)]}{P(X_1 = 1)} \cdot P(X_1 = 1) \\
&= \sum \pi_B \cdot b(x_1|x) + (1-\pi_B) \cdot \pi_C \cdot c(x_1|x) \\
&\quad + (1-\pi_0) \cdot [\lambda_B b(x_1|x) + (1-\lambda_B) \cdot \lambda_C \cdot c(x_1|x)] \\
&= \pi_B + (1-\pi_B) \cdot \pi_C + (1-\pi_0) \left[\cdot \lambda_B + (1-\lambda_B)\lambda_C \right] \\
&= \pi_0 + (1-\pi_0) \cdot p
\end{aligned} \tag{5.22}$$

The formula matches the results when we calculated $\mathbb{E}_x[\pi_1]$ directly.

5.3.2 Example: Calculating $Q^2(\pi)$ for Markov Chain Model With Three Regime

We show how to calculate Q for a Markov Chain with three regimes. We assume $X_1, \ldots$ are random variables governed by a Markov Chain distribution with $P(x_0 = 0) = 1$. The transition probability matrices are

$$A = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$$

and the two alternative regimes have the transition probabilities

$$B = \begin{pmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{pmatrix}$$

π_B	0.1
π_C	0.1
λ_B	0.4
λ_C	0.1
p	0.46

Table 5.6: Parameters for the Markov Chain model with three regimes

and

$$C = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

.

All other variables are defined in Table 5.6.

Our $Q(g(\pi))$ is defined as $\min\{g(\pi_B, \pi_C), c \cdot \pi + \mathbb{E}g(\pi_B, \pi_C)\}$ with $c = 1$. The definition for $g(\pi_B, \pi_C)$ is derived in the previous chapters:

$$g(\pi_B, \pi_C) = (1 - \pi) + I(\pi_B \geq \pi_C) \cdot \pi_C + I(\pi_B < \pi_C) \cdot \pi_B \tag{5.23}$$

For the calculations of Q we also need to calculate the probabilities for x_1. Our model has two possible states $\{0, 1\}$:

$$\begin{aligned} P(x_1 = 1 | x_0 = 0) &= \pi_B b(1|0) + (1 - \pi_B) \cdot \pi_C \cdot c(1|0) \\ &+ (1 - \pi_B)(1 - \pi_C) \cdot \lambda_B b(1|0) + (1 - \pi_B)(1 - \pi_C) \cdot (1 - \lambda_B)(\lambda_C) c(1|0) \\ &+ (1 - \pi_B)(1 - \pi_C) \cdot (1 - \lambda_B - \lambda_C + \lambda_B \lambda_C) \cdot a(1|0) \\ \\ &= 0.1 \cdot 0.2 + 0.9 \cdot 0.1 \cdot 0.5 \\ &+ 0.9 \cdot 0.9 \cdot 0.4 \cdot 0.2 + 0.9 \cdot 0.9 \cdot 0.6 \cdot 0.1 \cdot 0.5 \\ &+ 0.9 \cdot 0.9 \cdot 0.6 \cdot 0.9 \cdot 0.1 \\ &= 0.19784 \end{aligned} \tag{5.24}$$

and

$$\begin{aligned} P(x_1 = 0|x_0 = 0) = \pi_B b(0|0) + (1-\pi_B)\cdot \pi_C \cdot c(0|0) \\ (1-\pi_B)(1-\pi_C)\cdot \lambda_B b(0|0) + (1-\pi_B)(1-\pi_C)\cdot (1-\lambda_B)(\lambda_C) c(0|0) \\ (1-\pi_B)(1-\pi_C)\cdot (1-\lambda_B-\lambda_C+\lambda_B\lambda_C)\cdot a(0|0) \\ = 0.1\cdot 0.8 + 0.9\cdot 0.1\cdot 0.5 \\ + 0.9\cdot 0.9\cdot 0.4\cdot 0.8 + 0.9\cdot 0.9\cdot 0.6\cdot 0.1\cdot 0.5 \\ + 0.9\cdot 0.9\cdot 0.6\cdot 0.9\cdot 0.9 \\ = 0.80216. \end{aligned} \tag{5.25}$$

After calculating the conditional probabilities we are interested in calculating the conditional expectation. Meaning, we calculate π_1 for the cases $x_1 = 1$ and $x_1 = 0$. For further calculations, we start by calculating $\mathbb{E}[\pi_{\alpha,1}(0)]$.

$$\begin{aligned} \pi_{x,B,1}(0) &= \frac{\pi_B b(0|0) + (1-\pi_B)(1-\pi_C)\cdot \lambda_B b(0|0)}{numerator + (1-\pi_B)(1-\pi_C)\cdot (1-\lambda_B-\lambda_C+\lambda_B\lambda_C)\cdot a(0|0)} \\ &= \frac{0.3392}{0.80216} \\ &= 0.42289 \end{aligned} \tag{5.26}$$

and

$$\begin{aligned} \pi_{x,C,1}(0) &= \frac{(1-\pi_B)\cdot \pi_C \cdot c(0|0) + (1-\pi_B)(1-\pi_C)\cdot (1-\lambda_B)(\lambda_C) c(0|0)}{numerator + (1-\pi_B)(1-\pi_C)\cdot (1-\lambda_B-\lambda_C+\lambda_B\lambda_C)\cdot a(0|0)} \\ &= \frac{0.0693}{0.80192} \\ &= 0.08640. \end{aligned} \tag{5.27}$$

We can calculated $\mathbb{E}\pi_1(0)$ by adding both expected values:

$$\begin{aligned} \pi_{x,1}(0) &= \pi_{x,B,1}(0) + \pi_{C,1}(0) \\ &= 0.42289 + 0.08640 \\ &= 0.50930 \end{aligned} \tag{5.28}$$

Similar calculations resulted in:

$$\begin{aligned} \pi_1(1) &= 0.7789 \\ \pi_{B,1}(1) &= 0.4286 \\ \pi_{C,1}(1) &= 0.3503. \end{aligned} \tag{5.29}$$

After deriving the expected values and probabilities we can calculate $\mathbb{E}[\pi_1]$:

$$\begin{aligned} \mathbb{E}[\pi_1] &= \mathbb{E}[\pi_1(0)] \cdot P(x_1 = 0|x_0 = 0) + \mathbb{E}[\pi_1(1)] \cdot P(x_1 = 1|x_0 = 0) \\ &= 0.5093 \cdot 0.80192 + 0.7789 \cdot 0.19808 \\ &= 0.5627. \end{aligned} \tag{5.30}$$

We can check the value to the formula that we calculated previously

$$\begin{aligned} \mathbb{E}[\pi_1] &= \pi_B + (1 - \pi_B) \cdot \pi_C + (1 - (\pi_B + (1 - \pi_B) \cdot \pi_C)) \cdot p \\ &= 0.1 + 0.9 \cdot 0.1 + 0.81 \cdot 0.46 \\ &= 0.5626. \end{aligned} \tag{5.31}$$

Which is numerically equal to round of.

The goal of this section is to calculate $Q^1(\pi)$ or $g_1(\pi)$. The formula to derive this is given by

$$\begin{aligned}
g_1(\pi) &= \min\{(1-\pi) + I_{\{\pi_B \geq \pi_C\}}\pi_C + I_{\{\pi_B < \pi_C\}}\pi_B, c \cdot \pi \cdot \mathbb{E}[Q(\pi)]\} \\
&= \min\{1 - 0.2 + 0.1, 0.2 \cdot \mathbb{E}[Q(\pi)]\} \\
&= \min\{1 - 0.2 + 0.1, 0.2 \cdot 0.5759\} \\
&= \min\{0.9, 0.7759\} \\
&= 0.7759
\end{aligned} \tag{5.32}$$

with

$$\begin{aligned}
\mathbb{E}[Q(\pi)] &= [g(\mathbb{E}[\pi_{B,1}(0)], \mathbb{E}[\pi_{C,1}(0))]] \cdot P(x_1 = 0|x_0 = 0) \\
&\quad + [g(\mathbb{E}[\pi_{B,1}(1)], \mathbb{E}[\pi_{C,1}(1)]] \cdot P(x_1 = 1|x_0 = 0) \\
&= [(1 - 0.5093) + 0.0864] \cdot 0.80192 + [(1 - 0.7789) + 0.3503] \cdot 0.19784 \\
&= 0.4629 + 0.1130 \\
&= 0.5759.
\end{aligned} \tag{5.33}$$

5.3.3 Example: Results $Q^2(\pi)$ for Markov Chain Model With Three Regime

For the example in Section 5.3 we calculated the missing values. We will no be calculating $Q^2(\pi)$.

$$\begin{aligned}
g_2(\pi) &= \min\{g_1(\pi), \pi + \mathbb{E}[Q(\pi)]\} \\
&= \{0.8, 0.5 + \mathbb{E}[Q(\pi)]\}
\end{aligned} \tag{5.34}$$

For the expected value we can use the calculated values for $Q(\pi)$ by interpolating between the values.

$$\begin{aligned}
\mathbb{E}_\pi[g_1(\pi)] &= g_1(\pi) \cdot P(x_1 = 0|x_2 = 0) + g_1(\pi) \cdot P_\pi(x_1|x_0) \\
&= g_1(0.4868, 0.0768) \cdot P_\pi(x_1 = 0|x_2 = 0) + g_1(0.4927, 0.3110) \cdot P_\pi(x_1 = 1|x_0 = 0) \\
&= g_1(0.4868, 0.0768) \cdot 0.80216 + g_1(0.4927, 0.3110) \cdot 0.19784 \\
&= 0.5132 \cdot 0.80216 + 0.5075 \cdot 0.19784 \\
&= 0.51207.
\end{aligned} \tag{5.35}$$

This gives us

$$\begin{aligned}
g_2(\pi) &= \min\{0.3984, 0.3 + 0.51207\} \\
&= 0.3984.
\end{aligned} \tag{5.36}$$

6.

Simulations

We demonstrate the change point detection for a Markov chain model with three regimes using simulation and apply the calculated values of Q to our stopping rule.
We will use the same variables shown in Table 6.1 with different transition matrices.

6.1 Simulation 1

For the first simulation we are choosing an initial regime that makes a transition to state 1 very likely. Furthermore, the state 1 is an absorbing state, meaning that once entering state 1 the state does not change anymore. Therefore, a change point will be indicated immediately but not limited by a change from state 1 to state 0.

x_0	0
π_B	0.1
π_C	0.1
λ_B	0.4
λ_C	0.1
p	0.46
c	0.5
Q^n	Q^2

Table 6.1: Variables for the Markov Chain model with three regimes

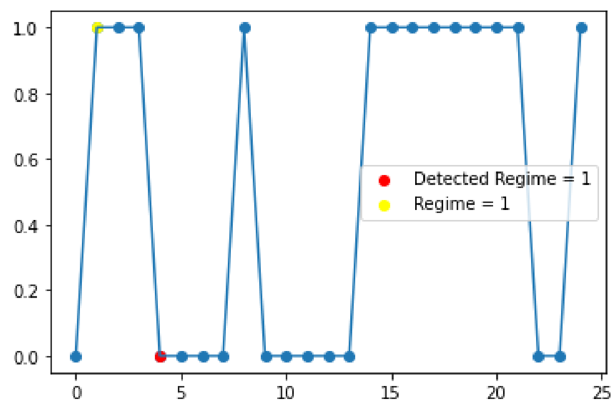


Figure 6.1: Simulation 1: Markov chain values

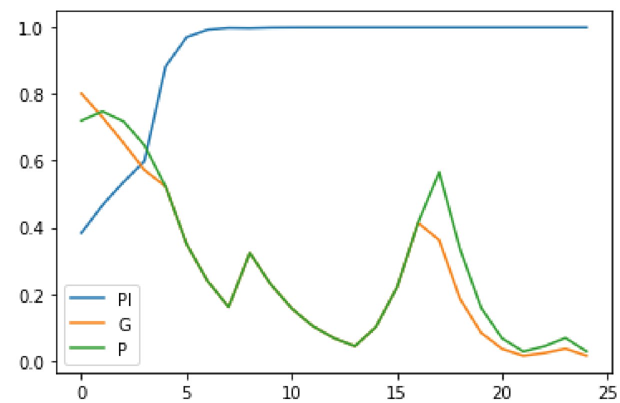


Figure 6.2: Simulation 1: π, $g(\pi_{n,B}, \pi_{n,C})$ and ρ

$$A = \begin{pmatrix} 0.3 & 0.7 \\ 0.0 & 1.0 \end{pmatrix}, B = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}, C = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

For the change point detection we have the following stopping rule:

$$\tau_0 = \inf\{n \geq 0 : \rho(\pi_{n,B}, \pi_{n,C}) \geq (1 - (\pi_{n,B} + \pi_{n,C})) + I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\}$$

where $\rho(\pi_{n,B}, \pi_{n,C})$ can be estimated by the calculated values of Q^n. We are using Q^{1000}. Therefore, the method will cause an alarm if

$$\tau_0 = \inf\{n \geq 0 : Q^{1000}(\pi_{n,B}, \pi_{n,C}) \geq (1 - (\pi_{n,B} + \pi_{n,C})) + I_{(\pi_{n,B} \geq \pi_{n,C})}\pi_{n,C} + I_{(\pi_{n,B} < \pi_{n,C})}\pi_{n,B}\}. \tag{6.1}$$

The first result is shown in Figure 6.1. The yellow point represents the actual change point while the red point stands for the caused alarm. As assumed, a change point is immediately detect if there is a switch from state 1 to state 0.

Furthermore the values of π_n, $g(\pi_{n,B}, \pi_{n,C})$ and $\rho(\pi_{n,B}, \pi_{n,C})$ are shown in Figure 6.2. As defined an alarm is caused when $\rho = g$.

6.2 Simulation 2

We are interested in seeing the same model in a second simulation to see how well the method detects a change point without an imediate change from state 1 to 0. The example of the change point detection can be seen in Figure 6.3. A switch happens at time step $t = 3$ (orange point) while an alarm is caused at $t = 24$. First, we could think that our change point detection is not optimal since a switch from state 1 to state 0 indicated that a change point has happened. However, the alarm is caused one time step after the switch. This can be explained by the additional component in our risk function: the cost of choosing the wrong regime. Even when it is certain that a change point has happened, the probability of choosing the wrong regime is too high.

An additional factor for the change point detection are the variables c_i. These are weights that can be adjusted to define the severity of early detection or detection delay just as choosing the wrong regime.

If we consider our Covid 19 data, these weights can be interpreted the following. Assume that we are prioritizing the health of the people and we want to detect changes as early as possible. Then we would weight the early detection factor much less. This will cause more early detections. However, the average detection delay will decrease. Another case is a business man who wants to avoid a lock down as much as possible. Here, an early detection can have a major impact on the person's business which is why they might weigh an early detection or the cost of choosing the wrong regime more.

6.3 Evaluation Simulation

We are interested in seeing how well our method detects change points. Therefore, we are using the model and simulate 100 data sets. Each time we are measuring the average early

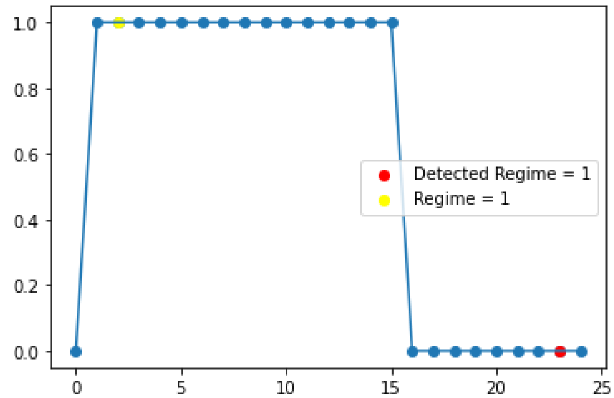


Figure 6.3: Simulation 2: Markov chain values

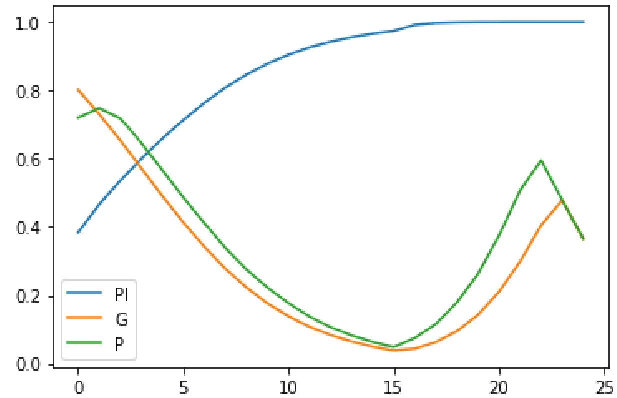


Figure 6.4: Simulation 2: π, $g(\pi_{n,B}, \pi_{n,C})$ and ρ

π	λ	A	B	C	Early Detection	Late Detection	Wrong Regime	c_1
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 0.5 & 0.5 \\ 0.9 & 0.1 \end{pmatrix}$	$B = \begin{pmatrix} 0.1 & 0.9 \\ 0.0 & 1.0 \end{pmatrix}$	$C = \begin{pmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{pmatrix}$	49 δ_e=3.32	51 $\delta_l = 0.83$	14 $\delta_B = 14$ $\delta_C = 0$	0.5
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}$	$B = \begin{pmatrix} 0.0 & 1.0 \\ 0.0 & 1.0 \end{pmatrix}$	$C = \begin{pmatrix} 0.0 & 1.0 \\ 1.0 & 0.0 \end{pmatrix}$	0.0 δ_e=0.0	98 $\delta_l = 1.0$ $\sigma_l = 0.0$	26 $\delta_B = 36$ $\delta_C = 0$	0.5
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$	$B = \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix}$	$C = \begin{pmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}$	54 δ_e=6.76	35 $\delta_l = 3.03$	1 $\delta_B = 1$ $\delta_C = 0$	0.5
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$	$B = \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix}$	$C = \begin{pmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}$	29 $\delta_e = 4.71$ $\sigma_e = 4.86$	71 $\delta_l = 5.69$ $\sigma_l = 3.52$	3 $\delta_B = 0$ $\delta_C = 0$	0.2
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$	$B = \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix}$	$C = \begin{pmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}$	12 δ_e=5.92 $\sigma_l = 5.71$	85 $\delta_l = 5.96$ $\sigma_l = 2.98$	1 $\delta_B = 1$ $\delta_C = 0$	0.05
(0.01,0.01)	(0.1,0.05)	$A = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$	$B = \begin{pmatrix} 0.25 & 0.75 \\ 0.25 & 0.75 \end{pmatrix}$	$C = \begin{pmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \end{pmatrix}$	32 $\delta_e = 2.77$	68 $\delta_l = 4.68$	14 $\delta_B = 0$ $\delta_C = 14$	0.2

Table 6.2: Results for a Markov chain model with 3 regimes. Bases on calculations with Q^2, $m = 11$ and 100 simulations

detection δ_e, the average late detection δ_l, and the number of regimes wrongly assigned. Here, δ_B names the cases where regime B was assigned while the actual regime was C. δ_C is applied analogously.

The results are shown in Table 6.2.

Every row shows a model with three regimes. In the first row, we have an initial regime with equal probabilities to stay or leave state 0 and a tendency to switch to state 0 once in state 1. The two alternative regimes are rather similar, both having a tendency towards state 1.

The initial probabilities for regime B or C are both 0.01. However, the parameter λ_B is 0.1

while λ_C is 0.05 and therefore greater than λ_C. This makes a transition to regime B more likely.

The results show an even distribution between early detection (49) and late detection (51) with an average early detection of 3.32 and an average detection delay of 0.83. 14 regimes were wrongly assigned to B when the regime was C. The weight for a detection delay was 0.5.

We are interested in a case where the transition probability matrices of B and C differ more significantly. We look at row 3 of our table Table 6.2. Here, the initial regime shows even probabilities of 0.5 for every transition. Regime B shows a clear tendency for state 1 while transitions to state 0 are very likely in regime C. The initial change point probabilities π and the variable λ stay the same as in row 1.

The results show that there are 54 early detections and 35 detection delays. Only one regime was assigned wrongly. However, we are interested in reducing the number of early detections and focus on causing an alarm when it actually happened. Therefore, we change the weight c_1 from 0.5 to 0.2.

Changing our weight results in less early detections and more detection delays. The average detection delay increased from 3.03 to 5.69 with a standard deviation of 3.52. The increase of the detection delay is due to the fact that early alarms are weighted more now. Therefore, the algorithm is rather careful about causing an alarm. In addition, the wrongly assigned regimes dropped from 1 to 0.

After changing lowering the value of c_1 to 0.05 (see row 5), we see an additional drop in early detection. The average detection delay only increased slightly while the standard deviation decreased. However, we see that one regime was assigned wrongly. This is most likely not due to the change of the weight.

The last row shows an example where the regimes B and C do not differ that drastically. We wanted to analyze how our method works when the transition probability matrices differ less significantly. We can see a slight increase in the detection delay. What is prominent that the wrongly assigned regimes increased by 14. All assigned C when the actual regime was B.

7.

Covid Data

We successfully developed a method to find a change point in Markov chain models with two alternative regimes. After testing our method on simulated data, we are interested in applying the approach on real time data. We start with the Wisconsin daily Covid data which is published on several websites such as Our World In Data. This chapter will be divided into four parts. First, we analyze the available data followed by developing a model for the Covid-19 data into a Markov chain model. The second section describes the result of our change point detection. Finally, we apply our results on other states data. Here, we use the derived change point detection and apply it on data from the state of Minnesota.

7.1 Wisconsin Covid Data

The Covid data is freely accessible at several websites. We used the data that is available on the "Our World In Data" website Data (2023). The first data we consider are the daily new confirmed Covid-19 cases. The daily new confirmed Covid-19 cases in Wisconsin are shown in Figure 7.1.

If we want to define our regimes, we would pick one regime where the cases are rather stable, meaning that there is no significant increase or decrease in the new confirmed cases. On the other hand we are interested in defining two other regimes. One showing a increase

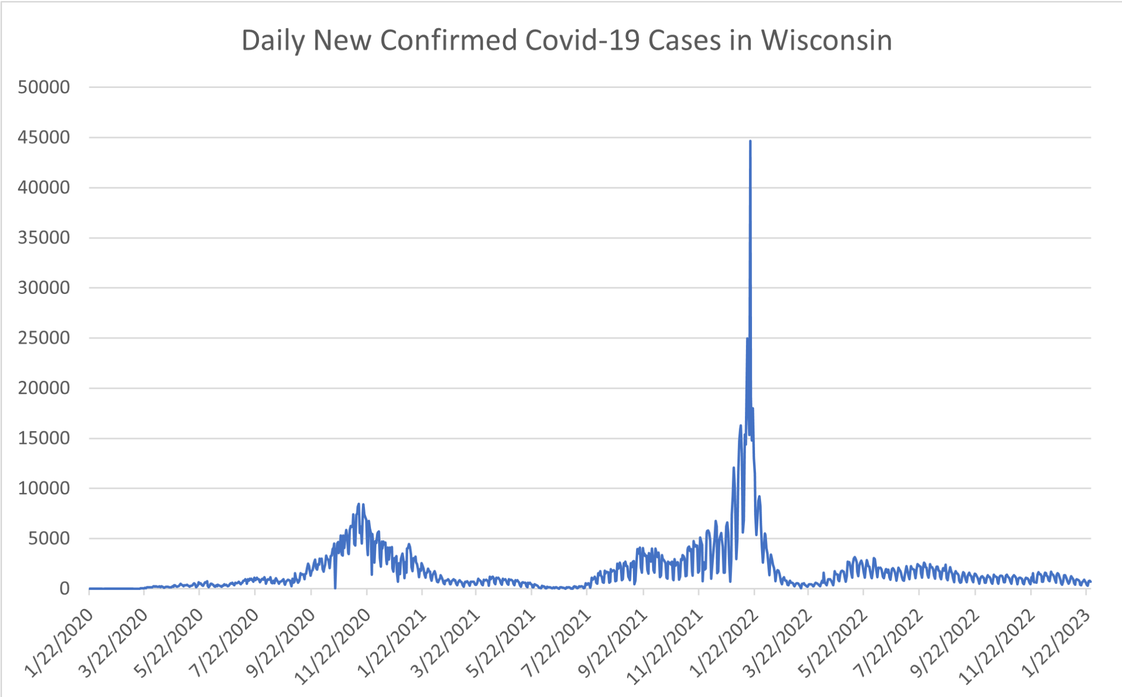


Figure 7.1: Daily New Confirmed Covid-19 Cases in Wisonsin. DHS (2023)

in new confirmed cases and the other showing a significant decrease in cases. We are labeling the regime with constant case numbers A, the regime with an increasing number of cases B and the last regime C. A first look at the data would suggest the regime A to take place in times such as January 2020 till June 2020 or May 2022 till November 2022. Examples for regime B could be August 2020 till October 2020, etc.

7.1.1 Model Wisconsin Data

Our simulations of the change point detection method are based on a Markov chain model with three states. Therefore, we establish a three state Markov chain for our Covid-19 data. One approach is to define the states through thresholds for the change in cases. The change can be calculated the following

$$\text{change}_{i+1} = n_{i+1} - n_i.$$

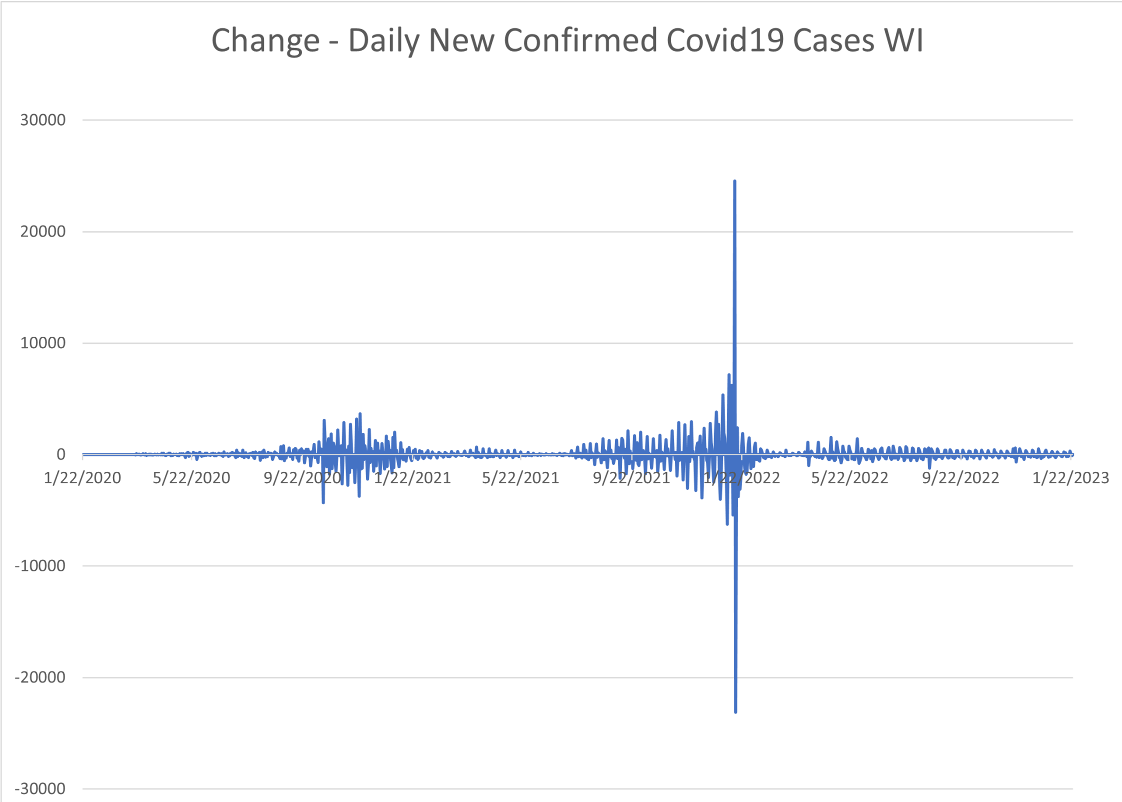


Figure 7.2: Change of Daily New Confirmed Covid-19 Cases in Wisonsin. DHS (2023)

State	0	1	2
change_{i+1}	≤ -70	$[-70, 70]$	≥ 70

Table 7.1: Thresholds for states of Covid-19 cases in Wisconsin

where n_{i+1} is the number of confirmed cases on day $i+1$ and n_i the number of confirmed cases on day i. The change in cases is shown in Figure 7.2 with an average change of 0.5259.

We are choosing the intervals for the different states given in Table 7.1.

With the chosen thresholds were able to categorize every change of Covid-19 cases into a state. The results are displayed in Figure 7.3. We can now use the data to estimate transition probabilities for every regime. Therefore, we need to define training data or intervals which can be used to estimate the transition matrices for regime A, B and C. We first choose to decide on the intervals by looking at the Daily New Confirmed Covid-19 cases (Figure 7.1). The data suggest that we are in regime A until August 2020 with a slow transition to

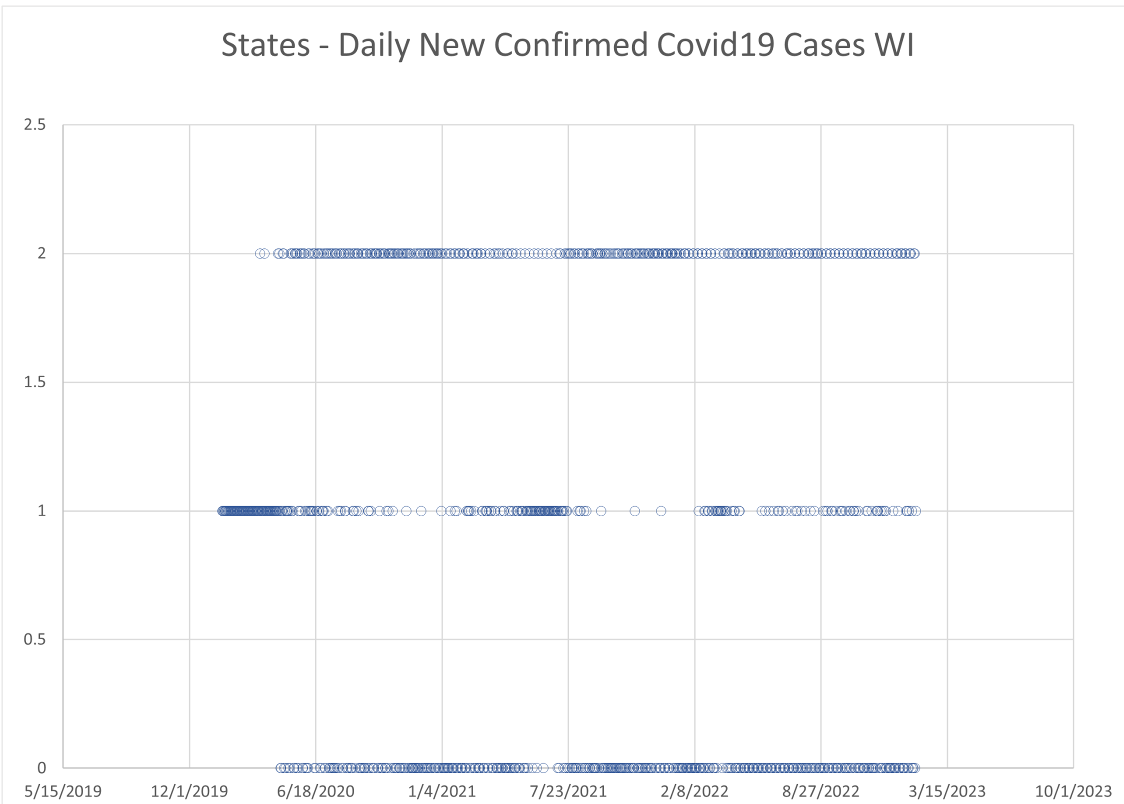


Figure 7.3: States of Daily New Confirmed Covid-19 Cases in Wisconsin. DHS (2023)

Regime	A	B	C
Training Interval	5/22/2020 – 8/22/2020	9/22/2020 – 10/22/2020	11/15/2020 – 1/15/2021

Table 7.2: Time intervals for estimating the transition probability matrices.

regime B. This increase in cases can be observed until October 2020. The definite decline can be seen starting Mid-November until Mid-January. We will try to use these intervals as training data. The exact intervals are shown in Table 7.2.

Each transition probability between state i and state j can be estimated by the following formula.

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \tag{7.1}$$

Using the training data gives us the transition probabilities

$$A = \begin{pmatrix} 0.3448 & 0.1379 & 0.5172 \\ 0.1154 & 0.3461 & 0.5385 \\ 0.4444 & 0.3611 & 0.1944 \end{pmatrix}$$

$$B = \begin{pmatrix} 0.3 & 0.1 & 0.6 \\ 0.5 & 0.0 & 0.5 \\ 0.3333 & 0.2 & 0.4667 \end{pmatrix}.$$

$$C = \begin{pmatrix} 0.4848 & 0.0 & 0.5152 \\ 0.5 & 0.0 & 0.5 \\ 0.68 & 0.08 & 0.24 \end{pmatrix}$$

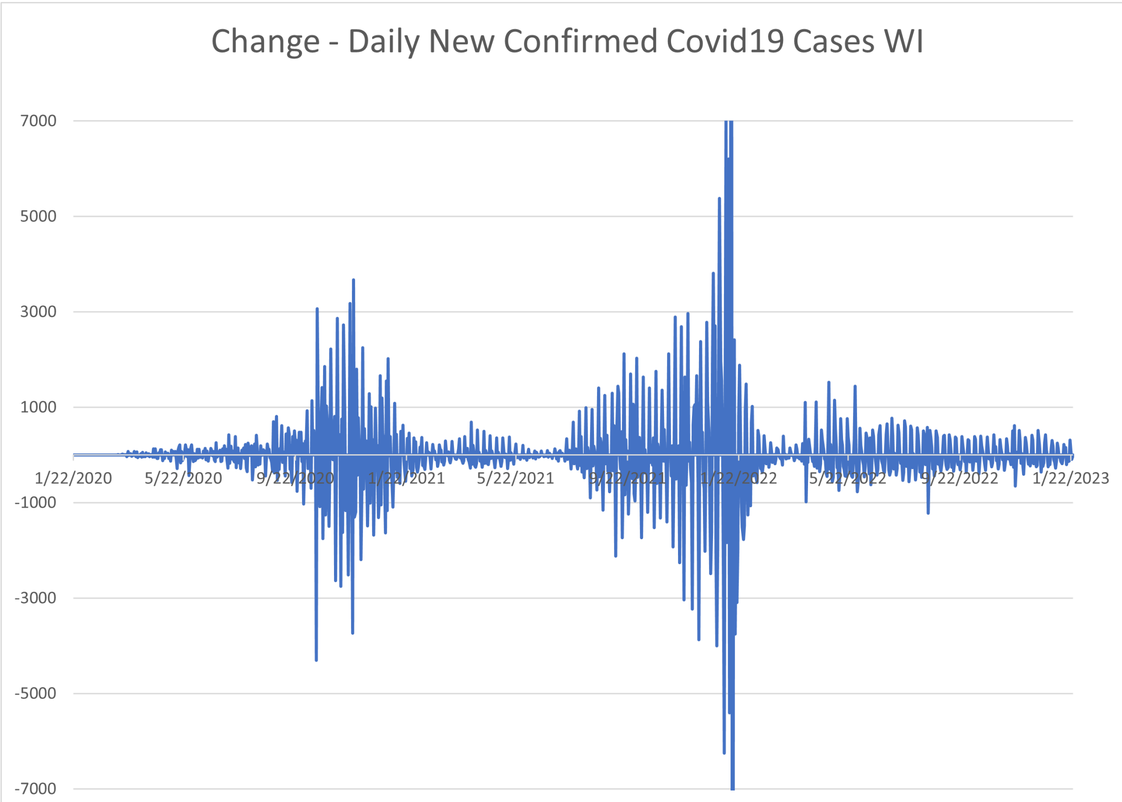


Figure 7.4: Change of Daily New Confirmed Covid-19 Cases in Wisonsin with adjusted range. DHS (2023)

Some of the transition probabilities are 0 which means that there were not any transitions between these states. This can be a result of having a small sample space. Therefore, we can either use more training data, by choosing more intervals for the parameter estimation. Another solution could be adjusting the intervals for the chosen states shown in Table 7.1. The sample size is relatively small since we only have Covid-19 data of the last two years. Dividing these data points into three different training sets for three regimes can be challenging. Therefore, we first try to adjust the thresholds for our states.

We take a closer look at the change in cases displayed in Figure 7.2. The extraordinary spike in change around January 2022 make it difficult to see the actual behavior pattern. Therefore, we are now looking at a graph were the range is going from -7000 to 7000 shown in Figure 7.4. The range is visible range is just omitted in the graph but not the actual data.

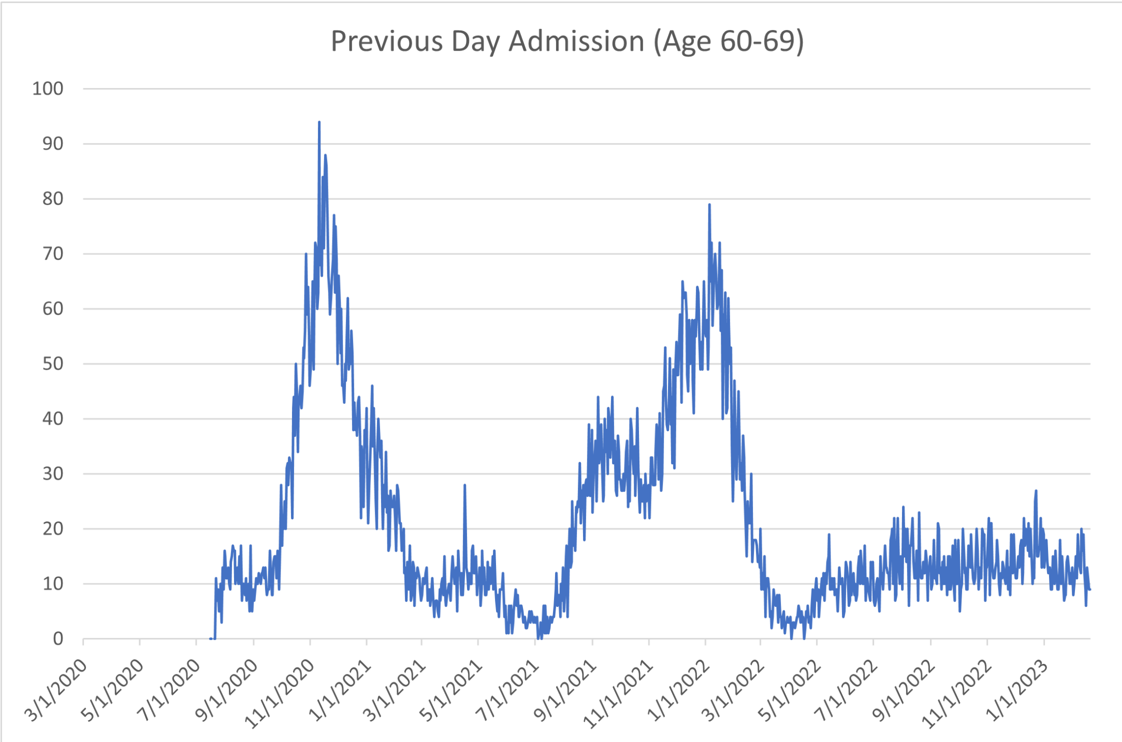


Figure 7.5: Previous Day Hospital Admissions (Age 60-69) in Wisconsin. HealthData.gov (2023)

The graph suggests that the thresholds -70 and 70 were making transitions to state 1 during an increase in cases rather unlikely.

Trying different intervals for the states does not result in complete transition probability matrices. Meaning, it is not possible to eliminate all zeros in the matrices.

7.2 Wisconsin Hospital Admission Data

Another data set that can indicate changes in the Covid-19 cases are the daily hospital admissions. The data is available on several websites such as HealthData.gov (2023). We choose to observe the daily hospital admissions for patients between 60 and 69. The numbers are shown in Figure 7.5.

We are again calculating the change between the current and previous day which results in the data shown in Figure 7.6.

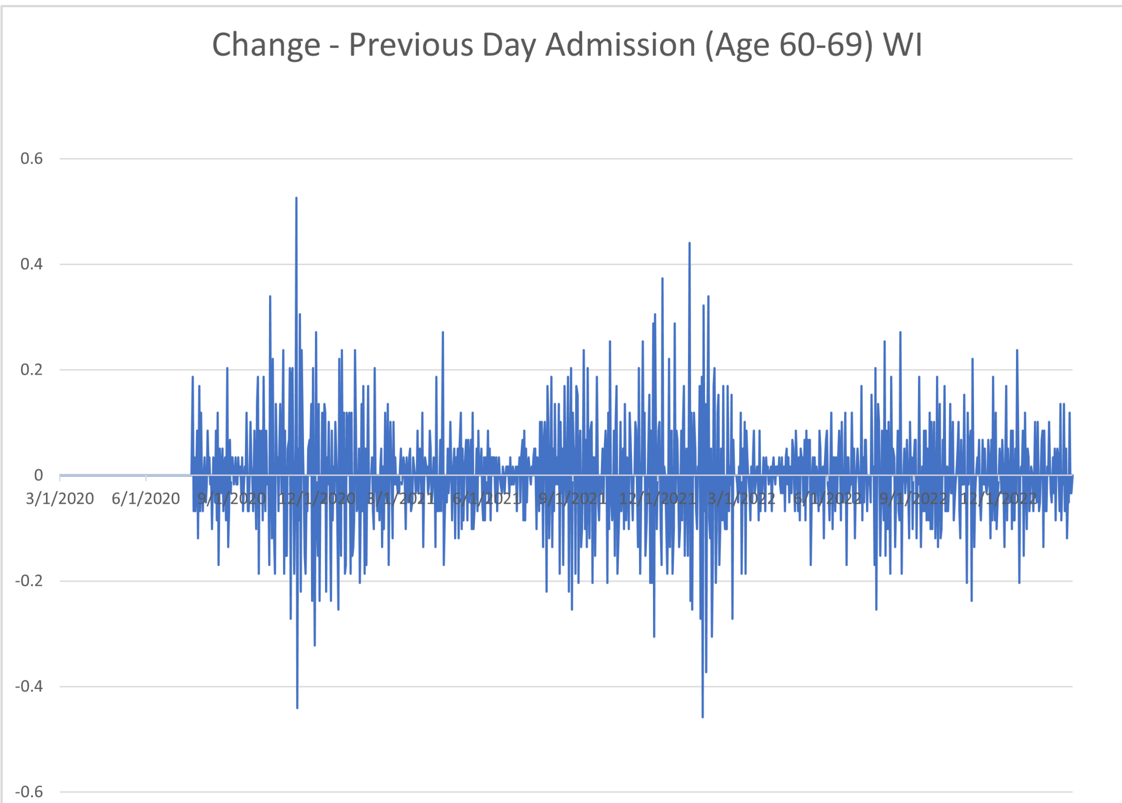


Figure 7.6: Change In Previous Day Hospital Admissions (Age 60-69) in Wisconsin. HealthData.gov (2023)

Regime	A	B	C
Training Interval	07/30/2020 – 09/22/2020 06/18/2022 – 12/01/2022	10/01/2020 – 11/06/2020 07/15/2021 – 09/11/2021	12/01/2020 – 02/10/2021 01/17/2022 – 03/15/2022

Table 7.3: Time intervals for estimating the transition probability matrices from Wisconsin hospital admissions.

State	0	1	2	3
change_{i+1}	≤ -0.1	$[-0.1, 0]$	$[0, 0.1]$	≥ 0.1

Table 7.4: Thresholds for states of Covid-19 Hospital admissions in Wisconsin

Again we are interested in organizing the changes into different states depending on the interval they are in. In order to have more training data, we are using several time intervals for the data. The chosen time frames for our training data are displayed in Table 7.3.

Next we are defining our states. In addition to choosing several time intervals, we are also using four states to categorize our changes in hospital admissions.

The resulting state distribution is shown in Figure 7.7.

With these setting we are able to calculate the transition probability matrices for our different regimes. We use the same method that is described previously. The results are three matrices that do not contain any probability equal to zero.

$$A = \begin{pmatrix} 0.0345 & 0.2759 & 0.4483 & 0.2414 \\ 0.0541 & 0.2703 & 0.4865 & 0.1891 \\ 0.1310 & 0.3929 & 0.3690 & 0.1071 \\ 0.4194 & 0.4194 & 0.1290 & 0.03226 \end{pmatrix}$$

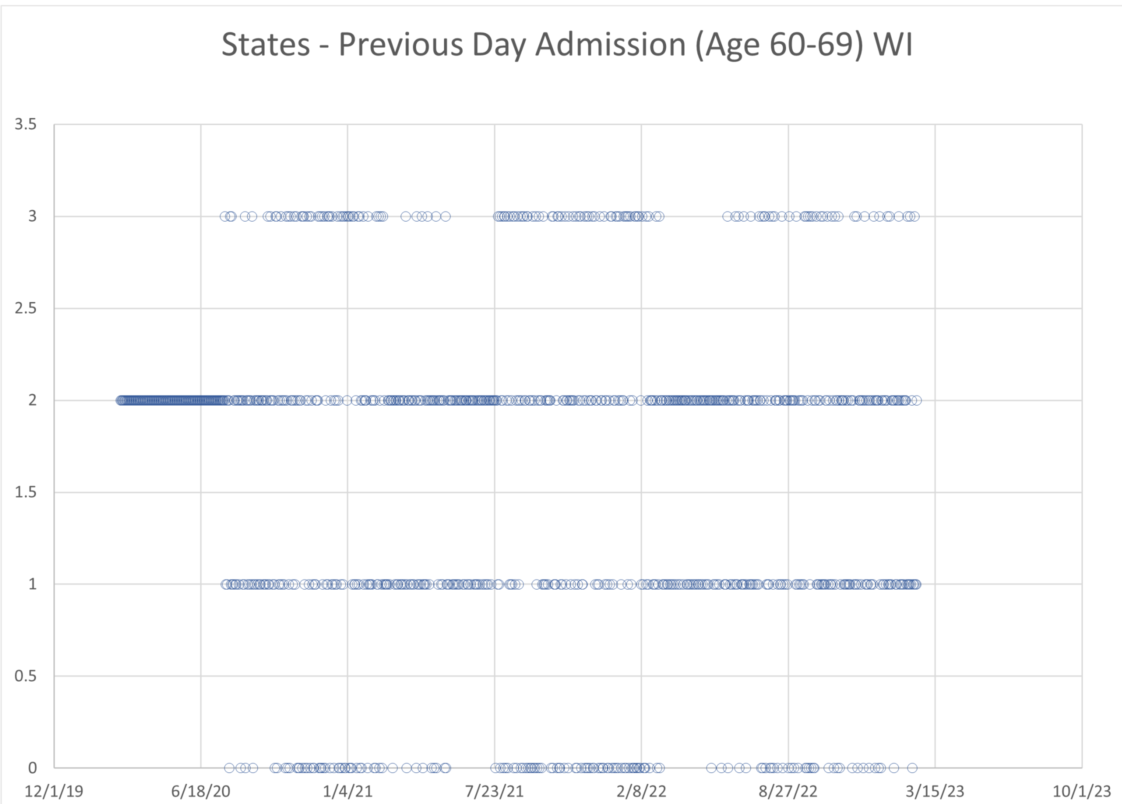


Figure 7.7: States In Previous Day Hospital Admissions (Age 60-69) in Wisconsin. HealthData.gov (2023)

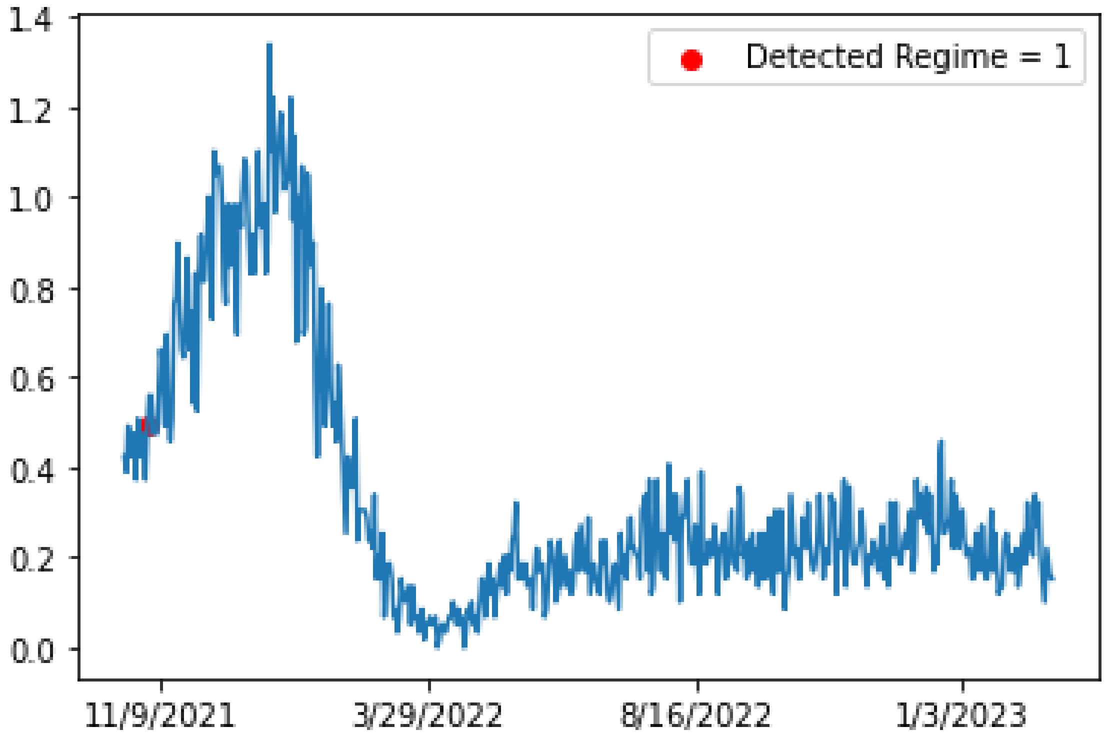


Figure 7.8: Change Point Detection on Wisconsin Hospital Data (Age 60-69) with start date on October 31st 2021

$$B = \begin{pmatrix} 0.1176 & 0.1176 & 0.3529 & 0.4118 \\ 0.125 & 0.0417 & 0.5417 & 0.2917 \\ 0.22 & 0.26 & 0.3 & 0.22 \\ 0.4571 & 0.2 & 0.2286 & 0.1143 \end{pmatrix}$$

$$C = \begin{pmatrix} 0.2581 & 0.0645 & 0.3548 & 0.3226 \\ 0.1818 & 0.1515 & 0.2727 & 0.3939 \\ 0.1667 & 0.4333 & 0.3 & 0.1 \\ 0.4063 & 0.375 & 0.0625 & 0.1563 \end{pmatrix}.$$

We can use these matrices of our regimes in our method to detect change points. There-

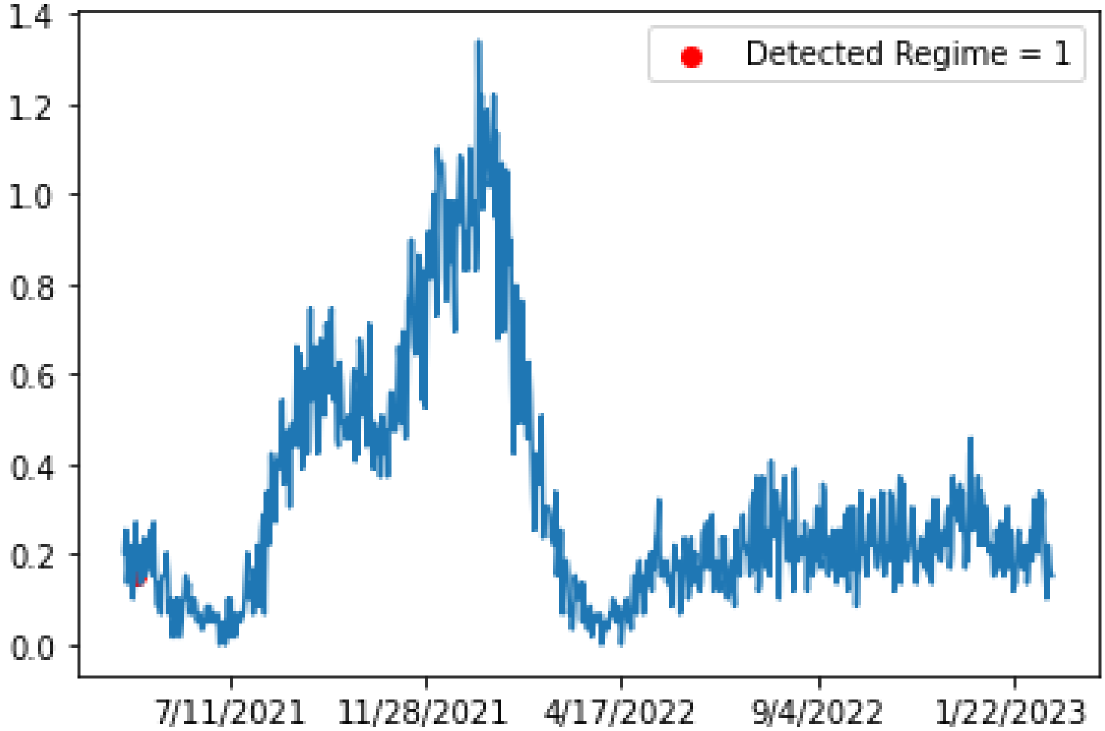


Figure 7.9: Change Point Detection on Wisconsin Hospital Data (Age 60-69) with start date on April 27th 2021

fore, we are looking back at the hospital admissions in Figure 7.5. To test our change point detection on real time data, we choose starting dates that were not used to estimate our probability matrices, for example 10/31/2021. The graph suggest that a week before the cases were increases with a short stabilization at the peek. We are now applying our change point detection assuming a initial regime A. The method detects a change point after 7 days with a switch to regime B as shown in Figure 7.8. In the graph regime 1 stands for regime B and regime 2 would be regime C.

While the change point and regime towards B seems to be detected correctly if we compare it to the data, the caused alarm seems to be early. While it was correct in this case, a very sensitive method can lead to false alarms in very small changes of behavior.

An example is looking at starting date April 27th 2021. The graph would suggest an initial regime A with a short decrease in cases and an increase after mid July. The current

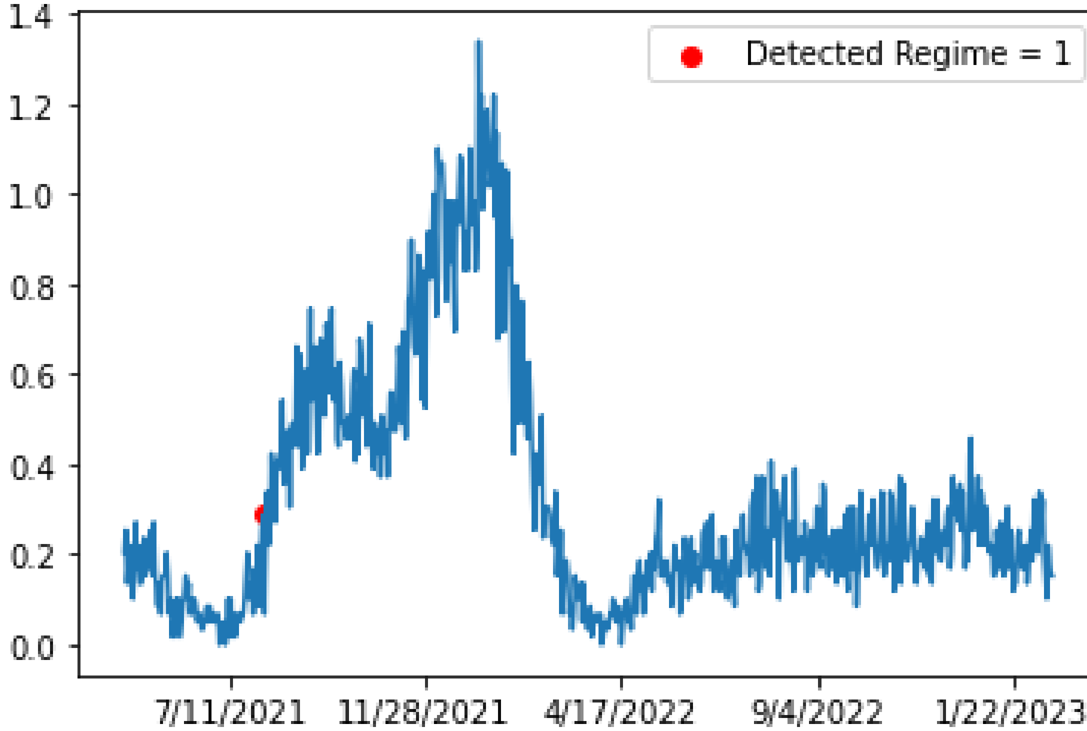


Figure 7.10: Change Point Detection on Wisconsin Hospital Data (Age 60-69) with start date on April 27th 2021 with $c_1 = 0.1$

method detects a change point after 7 days with a switch to regime B, shown in Figure 7.9. While there might be a short trend of increasing cases, the detection is too early.

A way to make the method less sensitive is adjusting the constant c_1 as we did in the simulation. Trying different constants $c_1 = 0.1$ results in good balance of early detection and late detection. With $c_1 = 0.1$ we detect a change point after 99 days towards regime B. The detection is shown in Figure 7.10.

While the method does not detect the short decrease in cases it detects the increase in hospital admissions quickly. This is due to the balance between early detection or predicting the wrong regime and the detection delay. This means that changing the constant resulted in a more accurate regime prediction but also into a detection delay and also a loss in sensitivity.

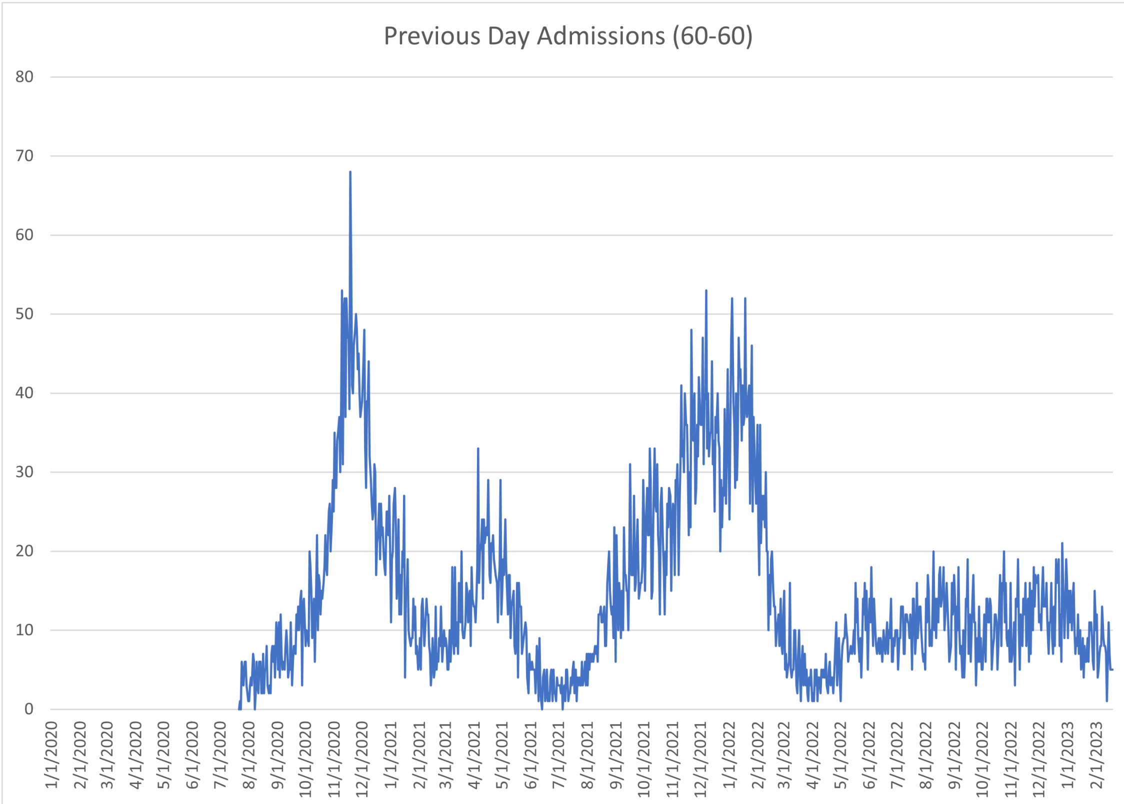


Figure 7.11: Previous Day Hospital Admissions (Age 60-69) in Minnesota. HealthData.gov (2023)

7.3 Change Point Detection On Other State Data

The Covid 19 daily data or the hospital admission data are very limited if we divide it into training data for three different regimes and additional testing data for our method. In the previous section we developed a change point detection method for Wisconsin hospital admission data and tested our method for different start dates. However, these start dates were very limited.

Therefore, we are testing our developed method with the calculated probability matrices on other states' data. We are choosing to test the method on the hospital admission data of Minnesota since it suggests to have a similar environment regarding factors such as population density and climate. The daily hospital admission data is shown in Figure 7.11.

First, we are choosing June 1st 2021 as our starting date. We decide for A as the initial

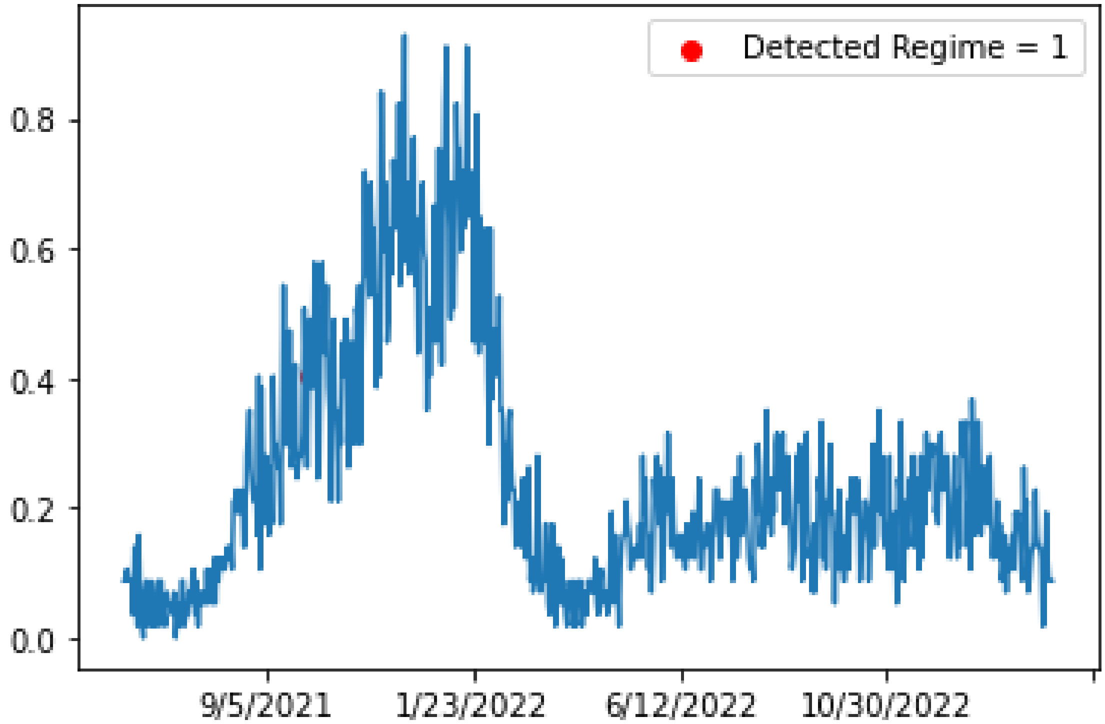


Figure 7.12: Change Point Detection on Minnesota Hospital Data (Age 60-69) with start date on June 1st 2021 with $c1 = 0.1$

regime. Looking at the development of the data we would assume a change point which results in an increasing numbers of cases. The detected change point is shown in Figure 7.12.

Another start date to choose from is November 20th 2022. In that time period the data suggest neither an increase nor a decrease in the data. Therefore, we can choose the regime A as our initial regime. We would expect a change point alarm in the next weeks towards regime C. The results of our change point detection method are shown in Figure 7.13.

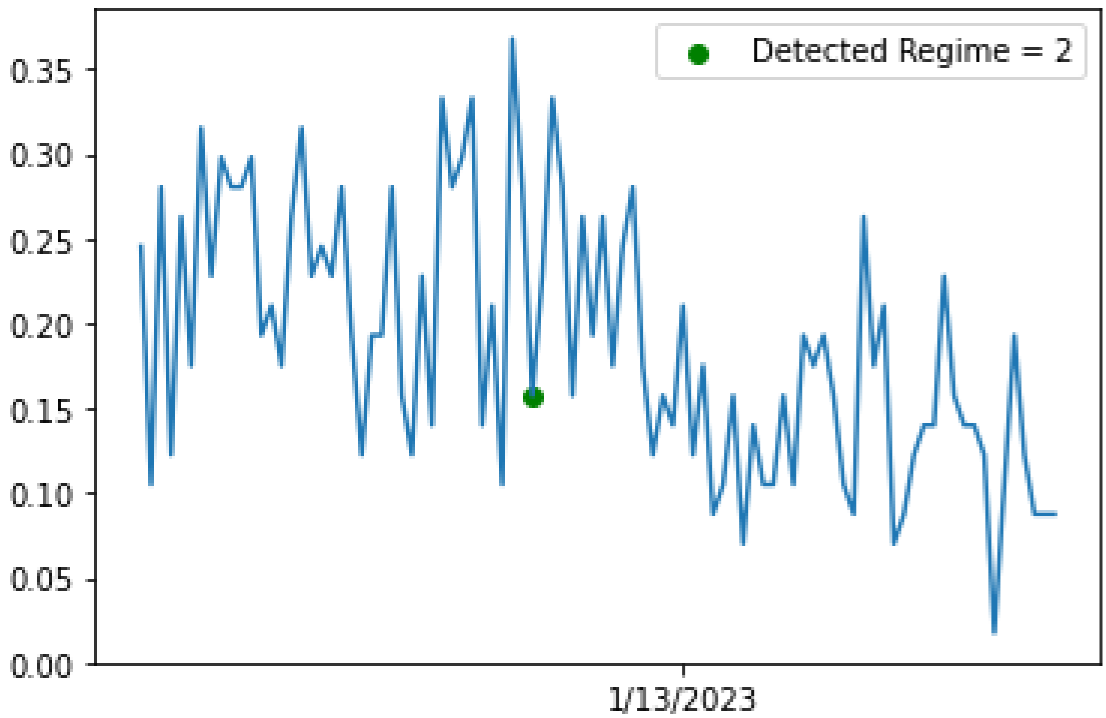


Figure 7.13: Change Point Detection on Minnesota Hospital Data (Age 60-69) with start date on November 20th 2022 with $c1 = 0.1$

8.

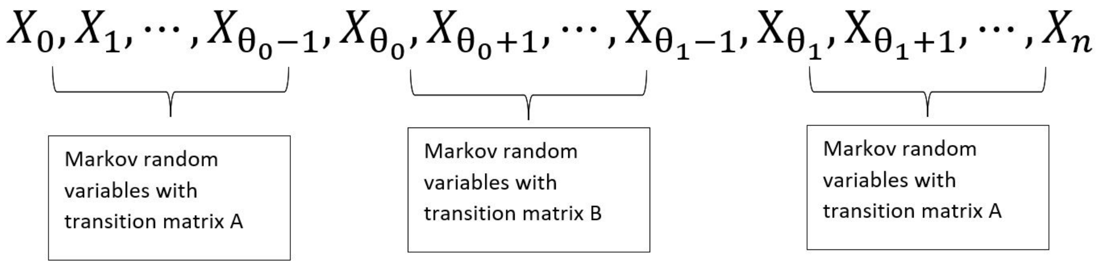
Summary

In this dissertation we developed a successful change point detection method for Markov chain models. The approach makes it possible to detect a change point and, additionally, decides on the new regime that the process had switched to.

We successfully applied and showed the functionality of the method on simulated Markov chain data. Later on, we were able to apply our method on Covid-19 data. Therefore, we transformed the collected Covid-19 hospital admission data into a Markov chain model with 3 regimes. The results showed that our method is also working on real time data can be modeled in Markov chain processes.

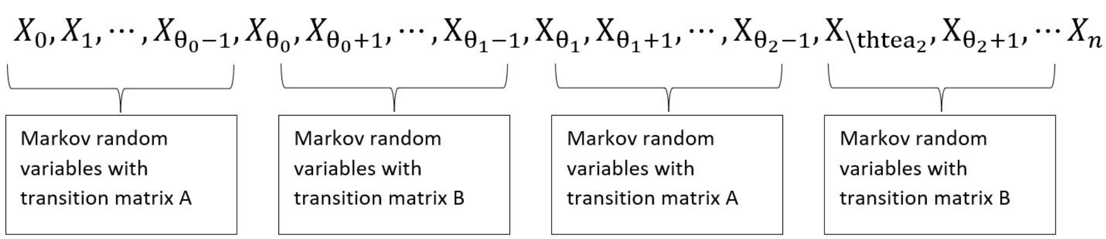
8.1 Multiple Change Point Detection With Two Regimes

An additional way to enhance the change point detection method is adding more change points. The random variables $X_0, X_1, \ldots, X_{\theta_0 - 1}$ would be governed by regime A. After the first change point θ_0 occurred the regime will switch to B, meaning that the random variables $X_{\theta_0}, X_{\theta_0+1}, \ldots, X_{\theta_1 - 1}$ have the transition matrix B. The second change point θ_1 will occur after change point θ_0 and will result in a switch to regime A.

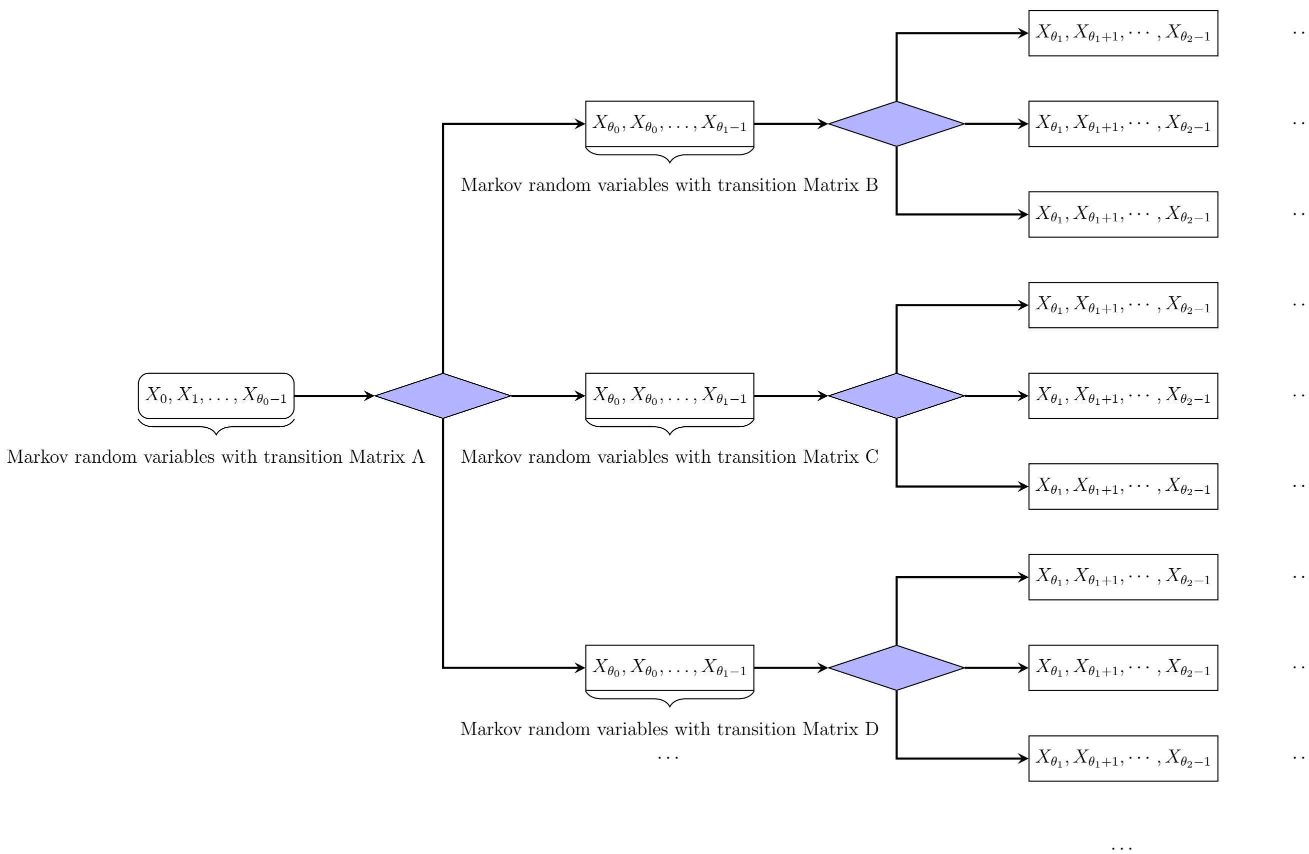


Again, we have to adjust the posterior probability and risk function.

After we would develop an approach for both, we could use the results to modify our approach to a multiple change point detection with two regimes. We will only have two possible regimes A and B that are switching when a change point θ_i occurs. We are getting $X_0, X_1, \ldots, X_{\theta_0-1}, X_{\theta_0}, X_{\theta_0+1}, \ldots, X_{\theta_1-1}, X_{\theta_1}, X_{\theta_1+1}, \ldots, X_{\theta_2-1}, X_{\theta_2}, X_{\theta_2+1}, \ldots$.



After finding an approach for the case of multiple change points and multiple regimes, the two method could be combined.



8.1.1 Change Point Detection With Multiple Regimes

After developing a change point detection method, we started to enhance our approach to a detection with multiple regimes. This section shows the beginning of this process, where the first Lemma are created. This can be a good foundation for enhancing the method.

We name our different regimes $r_0, r_1, \ldots, r_m$. The initial regime is r_0. The probability of switching to regime r_i at time θ_i is geometrically distributed with parameter λ_i. The change point θ is defined as the minimum of all θ_i's. The following lemma will derive the distribution of the change point θ.

Lemma 8.1. *Let $\theta_1, \ldots, \theta_m$ be independent, geometrically distributed i.i.d. random variables with parameters $\lambda_1, \ldots, \lambda_m$. Then the minimum of the random variables is also geometrically distributed with parameter $\lambda = \sum_{i=1}^{m} \lambda_i - \sum_{k=1}^{m} \sum_{j=1}^{i-1} \lambda_i \lambda_j + \sum_{i=1}^{m} \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \lambda_i \lambda_j \lambda_k \cdots - (-1)^m \prod_{l=1}^{m} \lambda_l$.*

Proof.

$P(\theta_i \geq k) = (1 - \lambda_i)^{k-1}$

$P(\min\{\theta_1, \ldots, \theta_m\}) = P(\theta_1 \geq k, \ldots, \theta_m \leq k)$

$= P(\theta_1 \geq k) \times \cdots \times P(\theta_m \geq k)$

$= (1 - \lambda_1)^k \times \cdots \times (1 - \lambda_m)^k$

$P(\min\{\theta_1, \ldots, \theta_m\} = k) = P(\min\{\theta_1, \ldots, \theta_m\} \geq k) - P(\min\{\theta_1, \ldots, \theta_m\} \geq k + 1)$

$= (1 - \lambda_1)^{k-1} \times \cdots \times (1 - \lambda_m)^{k-1} - (1 - \lambda_1)^k \times \cdots \times (1 - \lambda_m)^k$

$= (1 - \lambda_1)^{k-1} \times \cdots \times (1 - \lambda_m)^{k-1} - ((1 - \lambda_1) \times \cdots \times (1 - \lambda_m))(1 - \lambda_1)^{k-1} \times \cdots \times (1 - \lambda_m)^{k-1}$

$= (1 - ((1 - \lambda_1) \times \cdots \times (1 - \lambda_m)))(1 - \lambda_1)^{k-1} \times \cdots \times (1 - \lambda_m)^{k-1}$

$= (1 - ((1 - \lambda_1) \times \cdots \times (1 - \lambda_m)))((1 - \lambda_1) \times \cdots \times (1 - \lambda_m))^{k-1}$

$= (\lambda)(1 - \lambda)^{k-1}$

With

$$\lambda = 1 - \prod_{i=1}^{m}(1 - \lambda_i)$$

$$= 1 - (1 - \lambda_1) \times \cdots \times (1 - \lambda_m)$$

$$= 1 - (1 - \sum_{i=1}^{m} \lambda_i + \sum_{k=1}^{m}\sum_{j=1}^{i-1} \lambda_i\lambda_j - \sum_{i=1}^{m}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} \lambda_i\lambda_j\lambda_k \cdots + (-1)^m \prod_{l=1}^{m} \lambda_l)$$

$$= \sum_{i=1}^{m} \lambda_i - \sum_{k=1}^{m}\sum_{j=1}^{i-1} \lambda_i\lambda_j + \sum_{i=1}^{m}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} \lambda_i\lambda_j\lambda_k \cdots - (-1)^m \prod_{l=1}^{m} \lambda$$

□

Furthermore, we need to calculate the probability that we are switching to a regime r_i. Namely, we switch to regime r_1 if $P(\theta_1 \leq \theta_2, \ldots, \theta_1 \leq \theta_m)$, we switch to regime r_2 if $P(\theta_2 > \theta_1, \theta_2 \leq \theta_3, \ldots, \theta_2 \leq \theta_m)$ and so on.

8.2 Other Data

In this dissertation we successfully detected change points on real Covid-19 data. In addition, we were able to predict the regime that the process has switched to. However, the used data was limited to a time period of three years. After the process of training our Markov chain model and calculate the transition probability matrices, there was only a limited amount of testing data. Therefore, it would be interesting to see how the method can be applied on other real data that can be modeled in a Markov chain. Such data can consist of but is not limited to other epidemiological models, speech recognition or credit risks.Ribeiro (2023)

Bibliography

Data, O. W. I. (2023 (accessed January 30, 2023)). United states: Coronavirus pandemic country profile. `https://ourworldindata.org/coronavirus/country/united-states`.

DHS, W. (2023 (accessed January 30, 2023)). Covid-19 data by state boundary v2. `https://data.dhsgis.wi.gov/datasets/wi-dhs::covid-19-data-by-state-boundary-v2/about`.

Hao, N., Niu, Y. S., and Zhang, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist. Sinica*, 23(4):1553–1572.

HealthData.gov (2023 (accessed February 20, 2023)). Covid-19 reported patient impact and hospital capacity by state timeseries. `https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh`.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. IV*, pages 149–165. University of California Press, Berkeley-Los Angeles, Calif.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721.

L. Koepcke, G. Ashida, J. K. (2016). Single and mutliple change point detection in spike trains: Comparison of different cusum methods. *Frontiers in Systems Neuroscience*, 10(51).

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B*, 57(4):613–658. With discussion and a reply by the author.

Polansky, A. M. (2007). Detecting change-points in Markov chains. *Comput. Statist. Data Anal.*, 51(12):6013–6026.

Ribeiro, D. (2023 (accessed March 29th, 2023)). Markov chain applications in data science. `https://www.mysense.ai/2021/07/14/markov-chain-applications-in-data-science/`.

Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry. part i. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638):204–230.

Ross, R. and Hudson, H. P. (1917). An application of the theory of probabilities to the study of a priori pathometry.—part ii. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650):212–225.

Shiryaev, A. N. (2008). *Optimal stopping rules*, volume 8 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin. Translated from the 1976 Russian second edition by A. B. Aries, Reprint of the 1978 translation.

Yakir, B. (1994). Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-point problems (South Hadley, MA, 1992)*, volume 23 of *IMS Lecture Notes Monogr. Ser.*, pages 346–358. Inst. Math. Statist., Hayward, CA.