

# Parameter Estimation using the Tau-Leaping Algorithm

Bachelor Thesis Mathematics



# Universiteit Utrecht

Olaf de Graaff

Supervisor Dr. I. Kryven

18 June 2021

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Our model</b>	<b>4</b>
2.1	Reaction speed . . . . .	5
2.2	Stochastic simulation algorithm . . . . .	5
2.2.1	Implementation . . . . .	7
2.3	Tau-leaping algorithm . . . . .	8
2.3.1	Implementation . . . . .	9
<b>3</b>	<b>Parameter estimation given the number of occurred reactions</b>	<b>10</b>
3.1	Estimation of rate constant . . . . .	10
3.2	One dimensional Fisher information . . . . .	11
3.3	Confidence interval for a constant rate factor . . . . .	16
3.4	Estimation of linear rate constant . . . . .	17
3.5	Higher dimensional Fisher information . . . . .	17
3.6	Confidence interval linear rate factor . . . . .	18
<b>4</b>	<b>Parameter estimation given the concentrations</b>	<b>20</b>
4.1	Poisson difference distribution . . . . .	20
4.2	Maximum likelihood estimator for Poisson difference distribution . . . . .	23
4.3	Fisher information of Poisson difference distribution . . . . .	24
4.4	Confidence interval . . . . .	25
<b>5</b>	<b>Aggregation</b>	<b>26</b>
5.1	Confidence interval parameters . . . . .	28
5.2	Implementation . . . . .	29
<b>6</b>	<b>Percolation</b>	<b>31</b>
6.1	Confidence interval . . . . .	32
6.2	Goodness of fit . . . . .	33
<b>7</b>	<b>Discussion</b>	<b>35</b>
<b>8</b>	<b>Conclusion</b>	<b>36</b>

# Chapter 1

## Introduction

As the title already suggests, this paper will mainly focus on parameter estimation of certain kinds of reaction processes. A reaction can be seen as removing some balls from a Pólya urn and then replacing them with some new balls. In this case the balls can be anything, i.e. atoms, people, animals etc. The reactions that we will focus on, are reactions where the reaction speed is dependent on the concentration or number of the reactants. Chemical reactions mostly comply to this, since their reaction speed is dependent on the concentration of the reactants. Modelling a pandemic is also a good example, since the number of infections is dependent on the number of infectious and susceptible people. Especially in the last example, it is important that the reactions can be modelled correctly and that the parameters can be estimated accurately, since the values of the parameters are very important for determining which measures have to be taken to stop the pandemic.

In chapter 2 we will see how such reaction processes can be modelled. Many ways of how this can be done, are described in the article 'Modeling and Simulating Chemical Reactions' by Desmond J. Higham [1]. There are two algorithms that we will discuss; the stochastic simulation algorithm and the tau-leaping algorithm. In chapter 3 we will take a look at how we can estimate the values of the parameters of the processes using the idea of the tau-leaping algorithm. We will encounter well-known concepts from statistics, such as the maximum likelihood estimator and the Fisher information. In the chapter that follows we will derive the Poisson difference distribution, which is needed when we only measure the the change of concentration. This distribution is widely studied by Abdulhamid A. Alzaid and Maha A. Omair [2].

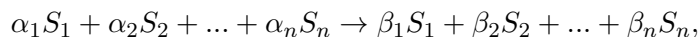
After that, we will look at aggregation processes, which is a special reaction process with some interesting properties. Lastly we will apply our theory on real data from two dimensional percolation and see if our model fits the reality.

All the Python code that is used in this paper can be found on Github, by the following link ([https://github.com/Olffie/Parameter\\_Estimation\\_using\\_the\\_Tau\\_Leaping\\_Algorithm](https://github.com/Olffie/Parameter_Estimation_using_the_Tau_Leaping_Algorithm)).

## Chapter 2

# Our model

The main inspiration for this chapter is the article 'Modeling and Simulating Chemical Reactions' written by Desmond J. Higham [1]. First, we will explain what type of model we will be looking at in this paper. Lets say, we have  $n$  different recognisable types. Those types can be anything, for example molecules, animal species or the current health state of individuals. Suppose we have  $n$  different kind of types. We keep track of the amount of each type by a state vector  $X(t)$ , that can change over time. The  $i$ 'th component of the state vector denotes the amount of the  $i$ 'th type. For types to change, we need reactions. The reactions will have the following form



in which  $\alpha$  and  $\beta$  are non-negative integers and  $S_i$  is type  $i$ .

To illustrate this, we will take an example from chemistry. Assume that we have four types of molecules,  $S_1, S_2, S_3$  and  $S_4$ . They can react with each other according through reactions of the following shape



When such reactions will take place, depends on many factors. Normally a reaction with multiple molecules will take place when these molecules collide with each other. So, theoretically we could keep track of the position and speed of each of the molecules and exactly model the whole system. This is called molecular dynamics. However, in practice this is quite hard for two reasons. First of all, it is almost impossible to determine the speed and place of each molecule exactly from its initial state, since molecules are extremely small. Secondly, it would take a lot of computing force to run such a model, because we have to calculate the position of every molecule at every small time step. Furthermore, it mostly does not tell us much about what is happening, but it only gives us knowledge about the outcome. That is why, this method is not commonly used in science. Instead we see the reactions as a probability process. In the coming sections and chapters we will use the reactions (2.1), (2.2) and (2.3) as an illustration for our theory.

## 2.1 Reaction speed

To run a model that is based on a probability process, we first need to know how many times a reaction will take place. If we take a closer look at reaction (2.1), we see what the reactants of the equation of  $S_1$  and  $S_2$  are. In chemistry, a reaction will take place if two molecules collide. That means that the reaction speed is proportional to the concentration of the reactants. If there are more molecules present, more reactions can take place. The reaction speed is known to be proportional to the product of the reactants. In other fields of science we can see the same phenomenon.

The reaction speed is, however, not only dependent on the concentrations of the reactants. Many different factors play a role, for example the speed, temperature, chance that a collision results in a reaction etc. To take these kinds of factors into account, we take every reaction to have a rate constant. Lets call the rate constant for reaction (2.1)  $c_1$ . In total, the reaction speed will be proportional to  $c_1 X_1 X_2$  with  $X_1$  and  $X_2$  denoting the concentrations of  $S_1$  and  $S_2$ . To keep our notation short, we will introduce a propensity function  $a_j(\mathbf{x})$ . The reaction speed of reaction  $j$  is proportional to  $a_j(\mathbf{x})$ , given a concentration (vector)  $\mathbf{x}$ . For reaction (2.1) we will have  $a_1(\mathbf{x}) = c_1 X_1 X_2$ .

Now we can give an expression for the probability that a reaction will take place in some interval. Take the interval  $[t, t + \tau)$  where  $\tau$  has to be small to such a degree that only a single reaction can take place in that interval. The probability that the first reaction will take place is then equal to  $a_1(x)\tau = c_1 X_1 X_2 \tau$ .

## 2.2 Stochastic simulation algorithm

We can now think of a simulation of our model. The first approach will be an exact algorithm, that is based on Desmond J. Highman's paper [1]. We will consider  $n$  different reactions with propensity functions  $a_j(\mathbf{x})$  for all  $j$  in  $\{1, \dots, n\}$ .

For deducing this algorithm, we first need to introduce a new probability function  $P(\tau | \mathbf{x}, t)$ , which is the probability that no reaction will take place in the time interval  $[t, t + \tau)$ . Since we have defined this function, we can determine what  $P(\tau + d\tau | \mathbf{x}, t)$  is, where we pick a sufficiently small value for  $d\tau$ , such that only one reaction can take place in the time interval  $[t + \tau, t + \tau + d\tau)$ .

Before we go any further, we make the assumption that  $P(\tau | \mathbf{x}, t)$  and  $P(d\tau | \mathbf{x}, t + \tau)$  are independent. Note that this is exactly true if no reactions took place in the interval  $[t, t + \tau)$  and approximately true when the reactions that are taking place in  $[t, t + \tau)$  are not changing the original concentration significantly. This is the case when  $\tau$  is small, since then the interval is relatively small which results in less reactions.

The assumption that  $d\tau$  is very small, such that there is a maximum of only one reaction in  $[t + \tau, t + \tau + d\tau)$ , can be used to express the probability that a reaction will take place in that time interval as  $d\tau \sum_{k=1}^n a_k(\mathbf{x})$ . We will write the summation over  $a$  as  $a_{sum}(\mathbf{x})$ . The law of total probability then gives us

$$P(d\tau | \mathbf{x}, t + \tau) = 1 - a_{sum}(\mathbf{x})d\tau.$$

Now we rewrite  $P(\tau + d\tau | \mathbf{x}, t)$  as

$$\begin{aligned} P(\tau + d\tau | \mathbf{x}, t) &= P(\tau | \mathbf{x}, t)P(d\tau | \mathbf{x}, t + \tau), \text{ (independent)} \\ &= P(\tau | \mathbf{x}, t)(1 - a_{sum}(\mathbf{x})d\tau). \end{aligned}$$

This expression can be rewritten in the following way:

$$\frac{P(\tau + d\tau | \mathbf{x}, t) - P(\tau | \mathbf{x}, t)}{d\tau} = -P(\tau | \mathbf{x}, t)a_{sum}(\mathbf{x}).$$

When we take the limit of  $d\tau \rightarrow 0$ , we end up with a linear ordinary differential equation

$$\frac{\delta P(\tau | \mathbf{x}, t)}{\delta \tau} = -P(\tau | \mathbf{x}, t)a_{sum}(\mathbf{x}).$$

Solving the differential equation gives us

$$P(\tau | \mathbf{x}, t) = P(0 | \mathbf{x}, t)e^{-a_{sum}(\mathbf{x})\tau} = e^{-a_{sum}(\mathbf{x})\tau}.$$

We now define  $R(\tau | \mathbf{x}, t)$  as the probability that a reaction took place in the interval  $[t, t + \tau)$ . Obviously we can see that,  $R(\tau | \mathbf{x}, t) = 1 - P(\tau | \mathbf{x}, t)$ . In this case  $R(\tau | \mathbf{x}, t)$  is nothing more than the cumulative distribution function of the reaction time of the first reaction. To find the probability density function, we have to take the derivative with respect to  $\tau$ . We get

$$r(\tau | \mathbf{x}, t) = a_{sum}(\mathbf{x})e^{-a_{sum}(\mathbf{x})\tau},$$

with  $r(\tau | \mathbf{x}, t)$  as the probability density function of the reaction time of the first reaction after time  $t$ .

We now know the probability of how long it will take before a reaction will take place. The only question that remains is, which reaction will take place. The chance of reaction  $i$  taking place as the first reaction is proportional to the propensity function, this means that the probability that the first reaction is reaction  $i$  equals  $a_i(\mathbf{x})/a_{sum}(\mathbf{x})$ . To summarise, we work with two random variables. One is continuous with an exponential distribution which defines the waiting time. The other is discrete and defines the reaction type.

Now we have enough information for deriving the algorithm. The algorithm consists of 6 steps.

- 1) Determine all values of  $a_i(\mathbf{x})$  and calculate  $a_{sum}(\mathbf{x})$ .
- 2) Draw two random independent standard uniform distributed numbers  $u_1, u_2$ .
- 3) Determine which reaction is going to take place. This can be done by finding the smallest  $j$  that satisfies  $\sum_{k=1}^j a_k(\mathbf{x}) \geq a_{sum}(\mathbf{x})u_1$ .
- 4) Now we want to know at what time the reaction is taking place. We can set  $R(\tau | \mathbf{x}, t) = u_2$  and solve for  $\tau$  to find how long it will take for a reaction to take place. Solving for  $\tau$  gives us  $\tau = -\log(1 - u_2)/a_{sum}$ .
- 5) Change the time  $t$  to  $t + \tau$  and change the state vector according to the reaction that took place.
- 6) Go back to step 1 or end the simulation.

### 2.2.1 Implementation

We can implement the stochastic simulation algorithm for the model that we have seen at the beginning of this chapter. For this model we use the following reaction rates as values

$c_1$	$c_2$	$c_3$
$2 \cdot 10^{-3}$	$4 \cdot 10^{-2}$	0.1

Table 2.1: Values of c

and as initial concentrations we take

$X_1$	$X_2$	$X_3$	$X_4$
500	200	0	0

Table 2.2: Initial values of X

Those initial values give us the graph that can be seen in figure 2.1.

From figure 2.1 we can clearly see that  $S_4$  never decreases. This is what we would expect, because nowhere in the equation is  $S_4$  a reactant. Another thing that can be observed is the fast increase of  $S_3$  at the beginning and then its decrease after  $t = 4$ . This phenomenon can also be explained. At the beginning, the first reaction will be dominant, because the other reactions are dependent on the concentration of  $S_3$ , which is low at the beginning of the process. However, after a certain amount of time the concentration of  $S_3$  is getting higher. This leads to an increase of reactions (2.2) and (2.3) wherein  $S_3$  is a reactant, which then causes a decrease of the concentration of  $S_3$ .

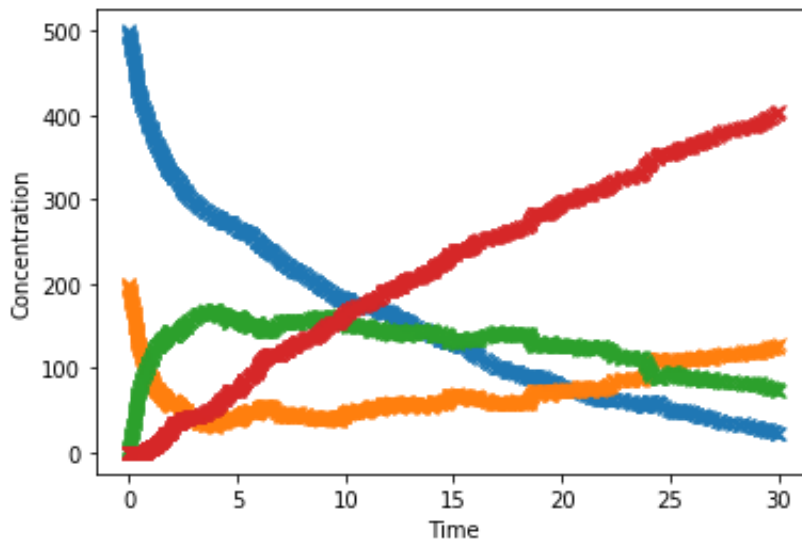


Figure 2.1: Implementation of the stochastic simulation algorithm,  
 $S_1$ : blue,  $S_2$ : orange,  $S_3$ : green,  $S_4$ : red.

## 2.3 Tau-leaping algorithm

There is one practical problem with the stochastic simulation algorithm and that is its slowness. After each reaction the propensity functions have to be recalculated and in each step you have to take two random variables. This is not a problem when there are only a few particles, however in scientific research these models can contain millions of particles. If this is the case the stochastic simulation algorithm is very time inefficient. That is why we are searching for a new algorithm for a new quicker algorithm that retains the accuracy of the stochastic simulation algorithm.

Instead of sampling  $\tau$ , we take small time steps of fixed size  $\tau$ , which we are free to define. We assume that in each time interval  $[t, t + \tau)$  the propensity functions will not change significantly. Now we can model how many times reaction  $i$  will take place within that time interval. We already know the probability of reaction  $i$  taking place in the interval  $[t, t + d\tau)$  where  $d\tau$  is infinitesimal, such that the reaction could at most only take place once, which is equal to  $a_i(\mathbf{x})d\tau$ . We now only have to count in how many of those infinitesimal time intervals of size  $d\tau$  in  $[t, t + \tau)$  the reaction occurred. This is known as a counting process. It is well-known that the number of reactions that take place in  $[t, t + \tau)$  will have a Poisson distribution with parameter  $a_i(\mathbf{x})d\tau$ . With this information we can construct the tau-leaping algorithm.

The tau-leaping algorithm consists out of 4 steps:

- 1) Determine  $a_i(\mathbf{x})$  for all  $i$ .
- 2) Draw samples from the  $Pois(a_i(\mathbf{x})\tau)$  distribution. This is the number of reactions of type  $i$  that occurred.
- 3) Change the state vector  $\mathbf{x}$  given the occurred reactions and update the time  $t$  to  $t + \tau$ .
- 4) Go back to step 1 or end.



The tau-leaping algorithm is much more time efficient than the stochastic simulation algorithm and therefore used much more often in scientific research. Thus, we will focus on the tau-leaping algorithm the next sections instead of the stochastic simulation algorithm.

### 2.3.1 Implementation

If we implement the tau-leaping algorithm in the model with the initial values from table 2.1 and 2.2, we get figure 2.2.

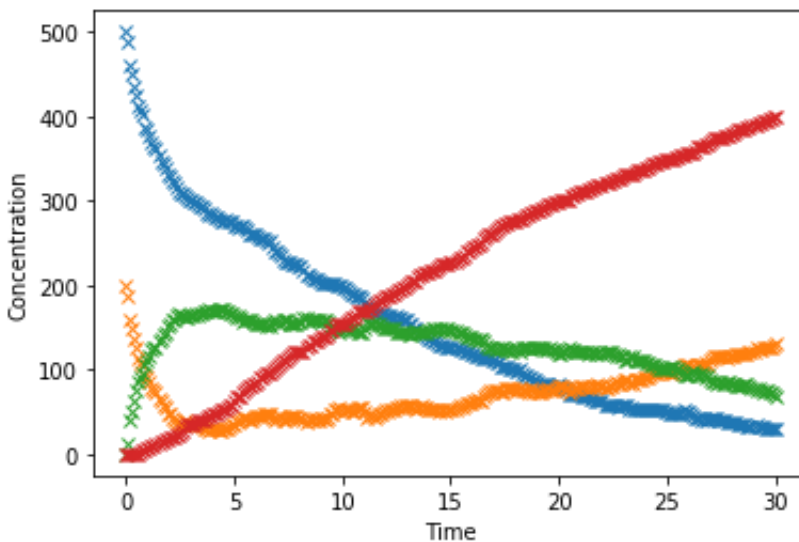


Figure 2.2: Implementation of the tau-leaping algorithm with  $\tau = 0.1$ ,  
 $S_1$ : blue,  $S_2$ : orange,  $S_3$ : green,  $S_4$ : red.

If we compare this figure with figure 2.1, we can see that we got approximately the same result. This gives an indication that the tau-leaping algorithm is almost as accurate as the stochastic simulation algorithm. It can never be as accurate as the stochastic simulation algorithm, since that is an exact algorithm. However, when we pick a proper value for  $\tau$ , it becomes quite close to an exact algorithm.

## Chapter 3

# Parameter estimation given the number of occurred reactions

In the previous chapter we saw how we can use different kinds of algorithms to model the system of reactions. For illustration purposes we will continue using the chemical model of particles.

Sometimes when you have a chemical model wherein particles react with each other, it is possible to measure the number of reactions that occurred and the consequent change in concentration. When using this procedure the rate constant remains unknown however. Nonetheless it is possible to give an estimation of the rate constant making use of statistics. In this chapter we will show how this can be done and how a confidence interval for the estimation can be made.

### 3.1 Estimation of rate constant

One of the most common ways of estimating parameters in statistics is with use of the maximum likelihood method. We assume that this method is well-known, if not the book ‘A Modern Introduction to Probability and Statistics’ can be consulted [3].

We will use the idea of the tau-leaping algorithm for estimating the rate constant, it is important that the time steps are sufficiently small such that the algorithm is an accurate representation of the real model.

Lets say we want to estimate  $c_1$ , which is the rate constant of (2.1). We assume that we know how many times this reaction took place in each time interval  $[i\tau, (i+1)\tau)$ . We define  $X_i(t)$  as the concentration of  $S_i$  at time  $t$ ,  $r_i$  as the number of reactions of type (2.1) that occurred in  $[i\tau, (i+1)\tau)$ , and  $R$  as  $[r_0, r_1, \dots, r_n]$ .

According to the tau-leaping algorithm, we know that  $r_i$  is  $Pois(X_1(t)X_2(t)c_1\tau)$  distributed. We define  $f_i$  as the probability mass function of the Poisson distribution at time step  $i$ . We can now try to find the maximum likelihood estimator. The likelihood is

$$L(c_1|R) = \prod_{i=0}^n f_i(r_i) = \prod_{i=0}^n \frac{(X_1(\tau i)X_2(\tau i)\tau c_1)^{r_i}}{r_i!} e^{-X_1(\tau i)X_2(\tau i)\tau c_1}. \quad (3.1)$$

To find the maximum likelihood estimator of  $c_1$  we must maximise this function. However, maximising this function directly is quite difficult. Therefore, we first take the logarithm of the likelihood. Since this new function will take the maximum value at the same point as the normal likelihood function, it is sufficient to just minimise the log-likelihood, which is

$$l(c_1|R) = \log(L(c_1|R)) = \sum_{i=0}^n r_i \log(X_1(\tau i)X_2(\tau i)\tau c_1) - \log(r_i!) - X_1(\tau i)X_2(\tau i)\tau c_1.$$

To maximise this function, we can take the derivative with respect to  $c_1$  and set it to be equal to zero. Doing so, gives us

$$\frac{\partial l(c_1|R)}{\partial c_1} = \sum_{i=0}^n \frac{r_i}{c_1} - X_1(\tau i)X_2(\tau i)\tau = 0, \quad (3.2)$$

so the maximum likelihood estimator of  $c_1$  then becomes

$$\hat{c}_1 = \frac{\sum_{i=0}^n r_i}{\sum_{i=0}^n X_1(\tau i)X_2(\tau i)\tau}. \quad (3.3)$$

For this value of  $\hat{c}_1$  we indeed have a minimum, since the second derivative is negative

$$\frac{\partial^2 l(c_1|R)}{\partial^2 c_1} = \sum_{i=0}^n -\frac{r_i}{c_1^2} < 0.$$

We can now start to debate the accuracy of this estimation. By making a 95%-confidence interval for  $c_1$  we can say more about the accuracy of the estimation. We can do this with the use of the Fisher information, before we do that however, we need to dive into the meaning and definition of the Fisher information.

## 3.2 One dimensional Fisher information

Most definitions and proofs in this section are based on the book ‘Mathematical Statistics and Data Analysis’ by John A. Rice. [4]. In this section we explore the properties of the Fisher information and how it can be used for making confidence intervals of estimated parameters. To do so, we first need a definition of the Fisher information.

**Definition 3.1.** Suppose we have a random variable  $X$  with a probability density function (or probability mass function) that is dependent on a variable  $\theta_0$ . We call this function  $f(X; \theta_0)$ . Then the Fisher information of variable  $\theta_0$  is defined as

$$\mathcal{I}(\theta_0) = E_{\theta_0} \left[ \left( \frac{\delta}{\delta \theta_0} \log f(X; \theta_0) \right)^2 \right].$$

This expression is, however, for the most part difficult to deal with. The following expression of the Fisher information is more commonly used.

**Lemma 3.1.** Suppose that  $f(X; \theta_0)$  is the probability density function of a continuous random variable  $X$ . If  $f(X; \theta_0)$  and  $\frac{\partial}{\partial \theta_0} f(X; \theta_0)$  are continuous on the sample space, then we can rewrite the Fisher information as

$$\mathcal{I}(\theta_0) = -E_{\theta_0} \left[ \frac{\delta^2}{\delta \theta_0^2} \log f(X; \theta_0) \right].$$

*Proof.* All the derivatives in this proof will be with respect to  $\theta_0$ . We start by taking the derivative of the log-likelihood. With use of the chain rule, we get

$$\frac{\delta}{\delta\theta_0} \log f(X; \theta_0) = \frac{f'(X; \theta_0)}{f(X; \theta_0)}.$$

In a similar fashion, we can rewrite the second derivative of the log-likelihood as

$$\begin{aligned} \frac{\delta^2}{\delta\theta_0^2} \log f(X; \theta_0) &= \frac{\delta}{\delta\theta_0} \left[ \frac{f'(X; \theta_0)}{f(X; \theta_0)} \right], \\ &= \frac{f''(X; \theta_0)f(X; \theta_0) - f'(X; \theta_0)^2}{f(X; \theta_0)^2}, \\ &= \frac{f''(X; \theta_0)}{f(X; \theta_0)} - \left( \frac{\delta}{\delta\theta_0} \log f(X; \theta_0) \right)^2. \end{aligned}$$

We now can take the expected value of the second derivative of the log-likelihood. Usage of the linearity and definition of the expected value gives us

$$\begin{aligned} -E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} \log f(X; \theta_0) \right] &= -E_{\theta_0} \left[ \frac{f''(X; \theta_0)}{f(X; \theta_0)} - \left( \frac{\delta}{\delta\theta_0} \log f(X; \theta_0) \right)^2 \right], \\ &= - \int \frac{f''(X; \theta_0)}{f(X; \theta_0)} f(X; \theta_0) dx + E_{\theta_0} \left[ \left( \frac{\delta}{\delta\theta_0} \log f(X; \theta_0) \right)^2 \right], \\ &= - \int f''(X; \theta_0) dx + \mathcal{I}(\theta_0). \end{aligned}$$

According to Leibniz integral rule, we can interchange the order of integration and differentiation, since  $f(X; \theta_0)$  and  $f'(X; \theta_0)$  are continuous. We get

$$\begin{aligned} &= - \frac{\delta^2}{\delta\theta_0^2} \int f(X; \theta_0) dx + \mathcal{I}(\theta_0), \\ &= - \frac{\delta^2}{\delta\theta_0^2} 1 + \mathcal{I}(\theta_0), \\ &= \mathcal{I}(\theta_0). \end{aligned}$$

□

Remark that the proof above also holds when we have a discrete random variable  $X$ . Therefore, the expression as in lemma 3.1 can also be used for Poisson random variables.

In this paper we will assume that all the Fisher information's that we encounter can be written in the form of lemma 3.1.

The Fisher information has also a property that is called the Chain rule, which we will need later in this section.

**Theorem 3.2** (Chain rule). *Suppose we have a sample  $\mathbf{X} = X_1, X_2, \dots, X_n$ , where all the  $X_i$  are independent, but not necessarily identical, distributed with probability density (or mass) function  $f_i(X_i; \theta_0)$ . The Fisher information is then given by*

$$\mathcal{I}_{\mathbf{X}}(\theta_0) = \sum_{i=1}^n \mathcal{I}_{X_i}(\theta_0).$$

*Proof.* Since the  $X_i$ 's are independent, we know that we can write the joint probability function as the product of the probability functions of the  $X_i$ 's. Using this fact and lemma 3.1 gives us

$$\begin{aligned} \mathcal{I}_{\mathbf{X}}(\theta_0) &= -E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} \log f(X_1, \dots, X_n; \theta_0) \right], \\ &= -E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} \log(f(X_1; \theta_0) \cdot \dots \cdot f(X_n; \theta_0)) \right], \\ &= -E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} (\log f(X_1; \theta_0) + \dots + \log f(X_n; \theta_0)) \right], \\ &= -E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} \log f(X_1; \theta_0) \right] - \dots - E_{\theta_0} \left[ \frac{\delta^2}{\delta\theta_0^2} \log f(X_n; \theta_0) \right], \\ &= \sum_{i=1}^n \mathcal{I}_{x_i}(\theta_0). \end{aligned}$$

□

One useful property of the Fisher information is that it can be used to construct a confidence interval for an estimated parameter. Before we start showing the correlation, we first state an important property of the maximum likelihood estimator. The following lemma is widely known for independent identically distributed samples, however it also holds for independent non-identically distributed samples [5].

**Lemma 3.3.** *Suppose that  $\mathbf{X} = X_1, X_2, \dots, X_n$  have probability density (or mass) function  $f_i(X_i, \theta_0)$  and that  $\hat{\theta}$  is the maximum likelihood estimator for  $\theta_0$  given the outcomes of  $X_1, X_2, \dots, X_n$ . Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \sigma_{\theta_0}),$$

*in other words the  $\sqrt{n}(\hat{\theta} - \theta_0)$  converges in distribution to a normal distribution with mean 0 and some variance  $\sigma_{\theta_0}$ .*

It turns out that the value of  $\sigma_{\theta_0}$  equals  $\mathcal{I}_{\mathbf{X}}(\theta_0)^{-1}$ . Which is stated in the theorem below. Mostly, this theorem is only stated for an independent distributed samples, however with some adjustments it can also be shown that it holds for an independent non identically distributed sample. This is important, since we have seen that we do not have identically distributed random variables when we apply the tau-leaping algorithm.

**Theorem 3.4.** *Suppose we have a sample  $\mathbf{X} = X_1, X_2, \dots, X_n$  of sample size  $n$ , where all the  $X_i$  are independent distributed with probability density function  $f_i(X_i; \theta_0)$  where  $\theta_0$  is an unknown parameter and  $\hat{\theta}$  as the maximum likelihood estimator of  $\theta_0$ . The asymptotic distribution of  $\hat{\theta}$  is then given as*

$$\mathcal{N}\left(\theta_0, \frac{1}{\mathcal{I}_{\mathbf{X}}(\theta_0)}\right).$$

*Proof.* We start again with defining  $l_n(\theta_0) = \sum_{i=1}^n \log f_i(X_i; \theta_0)$ . We find the value of the maximum likelihood estimator by solving  $l_n(\hat{\theta})' = 0$ . We can do a Taylor expansion at  $\theta_0$ .

We get  $0 = l_n'(\hat{\theta}) \approx l_n'(\theta_0) + (\hat{\theta} - \theta_0)l_n''(\theta_0)$ . This can be written as

$$\hat{\theta} - \theta_0 \approx \frac{-l_n'(\theta_0)}{l_n''(\theta_0)}.$$

Multiply each side with  $\sqrt{n}$ ,

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{-n^{-1/2}l_n'(\theta_0)}{n^{-1}l_n''(\theta_0)}. \quad (3.4)$$

Now we will derive the expected value and variance of the numerator,

$$E \left[ -n^{-1/2}l_n'(\theta_0) \right] = -n^{-1/2}E \left[ \sum_{i=1}^n \frac{\delta}{\delta\theta_0} \log f_i(X_i; \theta_0) \right].$$

If we now apply the linearity of the expected value and Leibniz integral rule, we end up with

$$-n^{-1/2} \sum_{i=1}^n \frac{\delta}{\delta\theta_0} E[f_i(X_i; \theta_0)] = -n^{-1/2} \sum_{i=1}^n 0 = 0.$$

For the variance we use the linearity, definition and Leibniz integral rule,

$$\begin{aligned} \text{Var} \left[ -n^{1/2}l_n'(\theta_0) \right] &= \frac{1}{n} \text{Var} \left[ \sum_{i=1}^n \frac{\delta}{\delta\theta_0} \log f(X_i; \theta_0) \right], \\ &= \frac{1}{n} \sum_{i=1}^n \text{Var} \left[ \frac{\delta}{\delta\theta_0} \log f(X_i; \theta_0) \right], \\ &= \frac{1}{n} \sum_{i=1}^n \left( E \left[ \left( \frac{\delta}{\delta\theta_0} \log f(X_i; \theta_0) \right)^2 \right] - E \left[ \frac{\delta}{\delta\theta_0} \log f(X_i; \theta_0) \right]^2 \right), \\ &= \frac{1}{n} \sum_{i=1}^n (\mathcal{I}_{X_i}(\theta_0) - 0), \\ &= \frac{1}{n} \mathcal{I}_{\mathbf{X}}(\theta_0). \end{aligned}$$

where we used the Chain rule (3.2) in the last step.

Now we take a look at the denominator of the fraction in equation (3.4). Since all the  $X_i$  are not necessarily identically independently distributed, we cannot apply the law of large numbers, because that is based on i.i.d. samples. However, there is a variation that is known as the Kolmogorov's strong law [6], which we can use in this case. Doing so gives us,

$$n^{-1}l_n''(\theta_0) = \frac{1}{n} \sum_{i=1}^n \frac{\delta^2}{\delta\theta_0^2} \log f_i(X_i; \theta_0) \xrightarrow{\text{a.s.}} E \left[ \frac{1}{n} \sum_{i=1}^n \frac{\delta^2}{\delta\theta_0^2} \log f(X_i; \theta_0) \right].$$

Using the linearity and the Chain rule (3.2) gives us,

$$E \left[ \frac{1}{n} \sum_{i=1}^n \frac{\delta^2}{\delta \theta^2} \log f(X_i; \theta_0) \right] = \frac{1}{n} \sum_{i=1}^n E \left[ \frac{\delta^2}{\delta \theta^2} \log f(X_i; \theta_0) \right] = \frac{1}{n} \sum_{i=1}^n -\mathcal{I}_{X_i}(\theta_0) = -\frac{1}{n} \mathcal{I}_{\mathbf{X}}(\theta_0).$$

This means that for a large  $n$  the expression  $\sqrt{n}(\hat{\theta} - \theta_0)$  can be written as

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{n^{1/2} l'(\theta_0)}{n^{-1} \mathcal{I}_{\mathbf{X}}(\theta_0)},$$

with expected value and variance equal to

$$E[\sqrt{n}(\hat{\theta} - \theta_0)] \approx 0,$$

$$Var[\sqrt{n}(\hat{\theta} - \theta_0)] \approx \frac{n^2}{\mathcal{I}_{\mathbf{X}}(\theta_0)^2} Var[-n^{1/2} l'(\theta_0)] = \frac{n^2}{\mathcal{I}_{\mathbf{X}}(\theta_0)^2} \frac{1}{n} \mathcal{I}_{\mathbf{X}}(\theta_0) = \frac{n}{\mathcal{I}_{\mathbf{X}}(\theta_0)}.$$

We already know that the expression  $\sqrt{n}(\hat{\theta} - \theta_0)$  is asymptotically normal distributed with mean zero and some variance (lemma 3.3). Above we have calculated what that variance is, so we get

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N} \left( 0, \frac{n}{\mathcal{I}_{\mathbf{X}}(\theta_0)} \right),$$

or in other words,

$$\hat{\theta} \sim \mathcal{N} \left( \theta_0, \frac{1}{\mathcal{I}_{\mathbf{X}}(\theta_0)} \right).$$

□

We now know how the maximum likelihood estimator is distributed and how the variance depends asymptotically on the Fisher information. There is only one problem with this distribution and that is its dependence on  $\theta_0$ . In practice, our goal is to estimate  $\theta_0$  by making use of the maximum likelihood estimator. If the distribution of  $\hat{\theta}$  is dependent on  $\theta_0$  we cannot determine how it is distributed. However, we know that the maximum likelihood estimator is a consistent estimator, so it converges in probability to  $\theta_0$ , in notation

$$\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0.$$

The Fisher information is a continuous function, so we can use the continuous mapping theorem. This gives us

$$\mathcal{I}(\hat{\theta}) \xrightarrow{\mathcal{P}} \mathcal{I}(\theta_0). \quad (3.5)$$

Combing (3.4) and expression (3.5), gives us the following result.

**Lemma 3.5.** *Suppose we have a sample  $\mathbf{X} = X_1, X_2, \dots, X_n$  of sample size  $n$ , where all the  $X_i$  are independent distributed with probability density function  $f_i(X_i; \theta_0)$  where  $\theta_0$  is an unknown parameter and  $\hat{\theta}$  as the maximum likelihood estimator of  $\theta_0$ . The asymptotic distribution of  $\hat{\theta}$  is then given as*

$$\mathcal{N} \left( \hat{\theta}, \frac{1}{\mathcal{I}_{\mathbf{X}}(\hat{\theta})} \right).$$

### 3.3 Confidence interval for a constant rate factor

In section 3.1 we have seen how we can determine the maximum likelihood estimator of the rate constant. With the knowledge of the previous section about the Fisher information, we can now also determine a confidence interval for our estimation. To do so, we first need to determine the Fisher information of our parameter  $c_1$ . We will do this by applying lemma 3.1. Doing this requires the second derivative of the log-likelihood. The first derivative has already been given in (3.2). When we take the derivative of (3.2), we get

$$\frac{\partial^2 l(c_1|X)}{\partial c_1^2} = -\frac{1}{c^2} \sum_{i=0}^n r_i.$$

Using lemma 3.1 gives us that the Fisher information of  $\hat{c}_1$  is equal to

$$\mathcal{I}_R(\hat{c}_1) = \frac{1}{\hat{c}_1^2} \sum_{i=0}^n r_i.$$

If we now assume that  $n$  is sufficiently large to use lemma 3.5, we get that the distribution of our estimated parameter  $\hat{c}_1$  is

$$\mathcal{N}\left(\hat{c}_1, \frac{\hat{c}_1^2}{\sum_{i=0}^n r_i}\right).$$

Which means that a  $100(1-\alpha)\%$  confidence interval for  $c_1$  can be written as

$$\left[ \hat{c}_1 - z(1 - \alpha/2) \sqrt{\frac{\hat{c}_1^2}{\sum_{i=0}^n r_i}}, \hat{c}_1 + z(1 - \alpha/2) \sqrt{\frac{\hat{c}_1^2}{\sum_{i=0}^n r_i}} \right], \quad (3.6)$$

with  $z(1 - \alpha/2)$  denoting the value  $x$  for which the probability  $P(Z < x) = 1 - \alpha/2$ , where  $Z$  is standard normal distributed.

In statistics we are mainly interested in a 95% confidence interval. We can construct such an interval for  $c_1$  with a data set. If we take 6 different data sets from our tau-leaping algorithm, with the initial values as in table 2.1 and 2.2,  $\tau = 0.1$  and  $t_{\text{final}} = 30$ , we get the following results rounded to 3 significant figures:

Data set	Lower bound	MLE	Upper bound	width
1	1.96	2.15	2.34	0.38
2	2.02	2.21	2.41	0.39
3	1.70	1.86	2.03	0.33
4	1.90	2.08	2.27	0.37
5	1.84	2.02	2.20	0.36
6	1.91	2.09	2.28	0.37

Table 3.1: 95% confidence intervals for  $c_1$ . All values should be multiplied by  $10^{-3}$ .



We can see that the width of the confidence intervals, which is a measure of the certainty, is mostly around  $3.6 \cdot 10^{-4}$ . In this case, we know the real value of  $c_1$ , which is  $2 \cdot 10^{-3}$ . Therefore we can conclude that we can estimate our  $c_1$  quite precisely without a large uncertainty.

### 3.4 Estimation of linear rate constant

In the previous sections we assumed that the rate constants did not change over time. Now we will assume that they are linear dependent on time, so  $c_i(t) = a_i t + b_i$ . In practice this can happen by increasing the temperature, since molecules will have a higher velocity and therefore will collide more. The methods that we are going to use in this section can also be applied to  $c_i(t)$ 's that have a higher order of  $n$ , however for the sake of simplicity we will continue using the linear dependency.

Lets take, for the first reaction (2.1) the rate constant  $c_1(t) = a_1 t + b_1$ . For estimating  $a_1$  and  $b_1$ , we can again apply the maximum likelihood method. Doing so eventually gives us the following two equations that we have to solve

$$\frac{\partial l(c_1|X)}{\partial a_1} = \sum_{i=0}^n \frac{r_i \cdot \tau i}{a_1 \tau i + b_1} - X_1(\tau i) X_2(\tau i) \tau^2 i = 0, \quad (3.7)$$

$$\frac{\partial l(c_1|X)}{\partial b_1} = \sum_{i=0}^n \frac{r_i}{a_1 \tau i + b_1} - X_1(\tau i) X_2(\tau i) \tau = 0. \quad (3.8)$$

Sometimes this can be solved by rewriting one of the equations and substituting it back in the other equation. That method, however, is not applicable on (3.7) and (3.8), since we cannot get the equations in a form wherein we can do a substitution. Therefore we have to solve this system of equations numerically. There are several methods for solving these equations, we will apply the Newton–Raphson method [7].

Solving these equations, does not give any information on the certainty of the estimated parameter. If we want to determine a confidence interval for  $a_1$  and  $b_1$  again, the one dimensional Fisher information, as described in section 3.2, is not sufficient anymore, since we have two estimated parameters that are dependent on each other. Nevertheless, we can compute a confidence interval of both parameters with the use of the Fisher information matrix, which is described in the next section.

### 3.5 Higher dimensional Fisher information

**Definition 3.2.** Suppose we want to estimate the parameter vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^T$ . The Fisher information matrix is then defined as

$$[\mathcal{I}(\boldsymbol{\theta})]_{i,j} = E_{\boldsymbol{\theta}} \left[ \left( \frac{\delta}{\delta \theta_i} \log f(X; \boldsymbol{\theta}) \right) \left( \frac{\delta}{\delta \theta_j} \log f(X; \boldsymbol{\theta}) \right) \right],$$

which can be rewritten again with the same conditions as in lemma 3.1 to

$$[\mathcal{I}(\boldsymbol{\theta})]_{i,j} = -E_{\boldsymbol{\theta}} \left[ \left( \frac{\delta}{\delta \theta_i \delta \theta_j} \log f(X; \boldsymbol{\theta}) \right) \right]. \quad (3.9)$$

For the higher dimensional Fisher information we also have a way of determining the distribution of  $\theta$ . In this case we have the following theorem, which is very similar with the one dimensional case [5].

**Theorem 3.6.** *Suppose we have a sample  $X_1, X_2, \dots, X_n$  of sample size  $n$ , where all the  $X_i$  are independent distributed with probability density function  $f(X; \theta)$  where  $\theta$  is an unknown parameter vector of dimension  $m$ . Denote  $\hat{\theta}$  as the maximum likelihood estimator of  $\theta$  given the data. The asymptotic distribution of  $\hat{\theta}$  is then*

$$\mathcal{N}\left(\hat{\theta}, [\mathcal{I}(\theta)]^{-1}\right),$$

where the  $-1$  indicates the inverse of the matrix.

### 3.6 Confidence interval linear rate factor

The Fisher information matrix can be used to make a confidence interval for the parameters  $a_1$  and  $b_1$ . In this case we have  $\theta = (a_1, b_1)$ , and the Fisher information matrix will be

$$\mathcal{I}(\theta) = \begin{bmatrix} \sum_{i=0}^n \frac{r_i \cdot (\tau i)^2}{(a_1 \tau i + b_1)^2} & \sum_{i=0}^n \frac{r_i \cdot \tau i}{(a_1 \tau i + b_1)^2} \\ \sum_{i=0}^n \frac{r_i \cdot \tau i}{(a_1 \tau i + b_1)^2} & \sum_{i=0}^n \frac{r_i}{(a_1 \tau i + b_1)^2} \end{bmatrix},$$

which followed from (3.7), (3.8) and (3.9). Now we can make a 95% confidence interval again by making use of theorem 3.6. We continue using the model that was described in (2.1)-(2.3). We take  $\tau = 0.1$ ,  $t_{\text{final}} = 30$ , the values listed in the table below,

$a_1$	$a_2$	$a_3$	$b_1$	$b_2$	$b_3$
$2 \cdot 10^{-4}$	$4 \cdot 10^{-3}$	0.01	$2 \cdot 10^{-3}$	$4 \cdot 10^{-2}$	0.1

Table 3.2: Values of  $c(t)$ .

and the initial concentration from table 2.2. If we determine a 95% confidence interval from 6 data sets from our code, we get the results rounded to 3 significant figures in table 3.3 and 3.4.

Data set	Lower bound	MLE	Upper bound	width
1	1.09	1.53	1.98	0.89
2	1.38	1.83	2.29	0.91
3	1.37	1.81	2.26	0.89
4	1.54	2.00	2.46	0.92
5	1.77	2.26	2.75	0.98
6	1.87	2.39	2.90	1.03

Table 3.3: 95% confidence intervals for  $a_1$ . All values should be multiplied by  $10^{-4}$ .

Data set	Lower bound	MLE	Upper bound	width
1	2.01	2.34	2.68	0.67
2	1.81	2.13	2.45	0.64
3	1.64	1.93	2.23	0.59
4	1.67	1.97	2.26	0.59
5	1.67	2.00	2.33	0.66
6	1.49	1.76	2.03	0.54

Table 3.4: 95% confidence intervals for  $b_1$ . All values should be multiplied by  $10^{-3}$ .

## Chapter 4

# Parameter estimation given the concentrations

In the previous chapter we discussed several ways of estimating the rate constant with the assumption that we know how many times each reactions occurred. This assumption, however, is not a realistic one, since in practice it is hard to count the number of reactions that occurred. It is much easier to measure the concentration at a certain time. If we take a look at our model from chapter 2, we see that we can sometimes determine the number of reactions that occurred by measuring the change in concentration. The reaction (2.3) is a perfect example of this. The trick is to measure the change in concentration of molecule  $S_4$ . Every extra molecule of  $S_4$  means one extra reaction of (2.3). So, for determining the rate constant, we can apply the same method as described in chapter 3.

We only encounter a problem when we want to determine the number of occurred reactions of (2.1) and (2.2). We cannot exactly determine how many times the reactions occurred, if we measure the concentrations of  $S_1, S_2$  or  $S_3$ . Let us call the number of reactions that occurred for (2.1)  $r_1$  and for (2.2)  $r_2$ . The only information that we could get by measuring  $S_1, S_2$  or  $S_3$  is the number  $r_1 - r_2$ . If we use the tau-leaping algorithm again, we know that the  $r_1 \sim Pois(X_1(t)X_2(t)c_1\tau)$  and  $r_2 \sim Pois(X_3(t)c_2\tau)$  in  $[t, t + \tau)$ . From which follows that  $r \sim Pois(X_1(t)X_2(t)c_1\tau) - Pois(X_3(t)c_2\tau)$ . The question that now arises is: “How can we estimate  $c_1$  and  $c_2$  given  $r$  over time?”. In this chapter we will show how this can be done. To do so, we first determine the distribution of the difference of two Poisson variables.

### 4.1 Poisson difference distribution

To determine the distribution of  $r$  we must know what the distribution is of the difference of two Poisson variables. This distribution is called the Poisson difference distribution, but is also known as the Skellam distribution. John Skellam was an English statistician who was the first person to derive the distribution of the Poisson difference distribution and therefore the distribution was named after him. The derivation of the distribution that we will provide in this section is based on his original proof [8].

The proof of the distribution of the Poisson difference distribution is based on probability generation functions. Before we start we will give a small recap of the definition and a theorem about the probability generation function. For the definition of a probability generating function we use the more generalised definition in which a random variable  $X$  can be negative as well. The following definition and lemmas can be found in ‘Probability generating functions for discrete real-valued random variables’ by M. L. Esquível [9].

**Definition 4.1.** Suppose  $X$  is a discrete random variable that can take the values  $\alpha_k$  with  $P(X = \alpha_k) = p_k$ . The probability generating function (PGF) of  $X$  is defined for all  $t > 0$  as

$$G_X(t) = E[t^X] = \sum_{k=-\infty}^{+\infty} p_k t^{\alpha_k}.$$

A fundamental theorem about the PGF is the following.

**Theorem 4.1** (Uniqueness of PGF). *Let  $X, Y$  be (discrete) random variables, such that  $G_X(t) = G_Y(t)$  for all  $t > 0$ . Then the probability distributions of  $X$  and  $Y$  are the same.*

Another important property of the PGF is stated in the following lemma.

**Lemma 4.2.** *Let  $X, Y$  be (discrete) random variables with PGF  $G_X(t)$  and  $G_Y(t)$ . The PGF of  $Z = X + Y$  is given as*

$$G_Z(t) = G_X(t)G_Y(t).$$

Lastly we need the following lemma before we can start by deriving the Poisson difference distribution. This lemma can be found in the ‘Handbook of Mathematical Functions’ by Milton Abramowitz and Irene A. Stegun [10].

**Lemma 4.3.** *Let  $t$  and  $z$  be real valued numbers ( $t \neq 0$ ), then we have that*

$$e^{\frac{z}{2}(t+1/t)} = \sum_{n=-\infty}^{\infty} t^n I_n(z),$$

where  $I_n(z)$  is the modified Bessel function of the first kind.

Now we can start with deriving the Poisson difference distribution, which is based on Skellam’s original proof [8].

**Theorem 4.4.** *Let  $X$  and  $Y$  be random Poisson variables with parameter  $\mu_1$  and  $\mu_2$  respectively and define  $Z = X - Y$ . The distribution of  $Z$  is known as the Poisson difference (or Skellam) distribution and equals*

$$P(Z = n) = e^{-(\mu_1 + \mu_2)} \left( \frac{\mu_1}{\mu_2} \right)^{n/2} I_n(2\sqrt{\mu_1 \mu_2}),$$

where

$$I_y(x) = \left( \frac{x}{2} \right)^y \sum_{k=0}^{\infty} \frac{(x^2/4)^k}{k!(y+k)!},$$

is known as the modified Bessel function of the first kind.

For  $Z$  we write that  $Z \sim PD(\mu_1, \mu_2)$ .

*Proof.* We start by determining the probability generation function of  $Z$ . This can be done with use of Lemma 4.2. Therefore, we first need to determine the PGF of  $X$  and  $-Y$ . We get that

$$\begin{aligned} G_X(t) &= \sum_{k=0}^{\infty} \frac{e^{-\mu_1} \mu_1^k}{k!} t^k, \\ &= e^{-\mu_1} \sum_{k=0}^{\infty} \frac{(\mu_1 t)^k}{k!}, \\ &= e^{-\mu_1} e^{\mu_1 t}, \\ &= e^{\mu_1(t-1)}, \end{aligned}$$

$$\begin{aligned} G_{-Y}(t) &= \sum_{k=0}^{\infty} \frac{e^{-\mu_2} \mu_2^k}{k!} t^{-k}, \\ &= e^{-\mu_2} \sum_{k=0}^{\infty} \frac{(\mu_2 t^{-1})^k}{k!}, \\ &= e^{-\mu_2} e^{\mu_2 t^{-1}}, \\ &= e^{\mu_2(t^{-1}-1)}. \end{aligned}$$

With use of lemma 4.2 we get that

$$G_Z(t) = G_X(t)G_{-Y}(t) = e^{\mu_1(t-1)}e^{\mu_2(t^{-1}-1)} = e^{-(\mu_1+\mu_2)+\mu_1 t+\mu_2 t^{-1}}. \quad (4.1)$$

Now we need to work back to a probability mass function, which can be done by making use of a smart substitution. We do the substitution  $\gamma = 2\sqrt{\mu_1\mu_2}$  and  $\phi = \sqrt{\frac{\mu_1}{\mu_2}}$ . Some algebraic steps and substitutions will lead to the expressions  $\mu_1 = \frac{\gamma}{2}\phi$  and  $\mu_2 = \frac{\gamma}{2}\phi^{-1}$ .

If we substitute these expressions in (4.1), we get that

$$\begin{aligned} G_Z(t) &= \exp \left\{ -\frac{\gamma}{2}(\phi + \phi^{-1}) + \frac{\gamma t}{2}\phi + \frac{\gamma}{2}(\phi t)^{-1} \right\}, \\ &= \exp \left\{ -\frac{\gamma}{2}(\phi + \phi^{-1}) + \frac{\gamma}{2}(t\phi + (\phi t)^{-1}) \right\}, \\ &= \exp \left\{ -\frac{\gamma}{2}(\phi + \phi^{-1}) \right\} \exp \left\{ \frac{\gamma}{2}(t\phi + (\phi t)^{-1}) \right\}. \end{aligned}$$

Which can be rewritten with use of lemma 4.3 as

$$\begin{aligned} G_Z(t) &= \exp \left\{ -\frac{\gamma}{2}(\phi + \phi^{-1}) \right\} \sum_{n=-\infty}^{\infty} t^n \phi^n I_n(\gamma), \\ &= \sum_{n=-\infty}^{\infty} \exp \left\{ -\frac{\gamma}{2}(\phi + \phi^{-1}) \right\} t^n \phi^n I_n(\gamma). \end{aligned}$$

If we compare this expression to the definition of a PGF in 4.1. We can conclude that the probability  $P(Z = \alpha_z) = p_z$  equals

$$P(Z = z) = \exp\left\{-\frac{\gamma}{2}(\phi + \phi^{-1})\right\} \phi^z I_z(\gamma).$$

Substituting this back in terms of  $\mu_1$  and  $\mu_2$  gives us

$$p_z(z) = e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2}\right)^{\frac{z}{2}} I_z(2\sqrt{\mu_1\mu_2}),$$

which is the probability mass function of the Poisson difference distribution. □

## 4.2 Maximum likelihood estimator for Poisson difference distribution

We have seen what the probability mass function of the difference of two Poisson variables is. Now we can return to our original problem wherein we are trying to estimate a rate constant. We will again look at (2.1) with the assumption that we can only measure the change in concentration. In other words, we only know the outcome of the random variable  $r$ , which was described at the beginning of this chapter. In this case we have, that  $r \sim PD(X_1(t)X_2(t)c_1\tau, X_3(t)c_2\tau)$ . Abdulhamid A. Alzaid and Maha A. Omair showed that you could determine the parameters of a Poisson difference distribution given an i.i.d. sample [2]. The method they used is only applicable to an i.i.d. sample, therefore we need to make some adjustments in their method before we can apply it. How this can be done for our model, is shown below.

We use the normal procedure to determine the maximum likelihood estimator of  $c_1$ . We assume that we have an independent sample  $R = [r_0, r_1, \dots, r_n]$ , where every  $r_i$  is Poisson difference distributed with parameter  $X_1(\tau i)X_2(\tau i)c_1\tau$  and  $X_3(\tau i)c_2\tau$  respectively. To keep the notation short, we define  $\theta_1(t) = X_1(t)X_2(t)\tau$  and  $\theta_2(t) = X_3(t)\tau$ , so we have  $r_i \sim PD(\theta_1(\tau i)c_1, \theta_2(\tau i)c_2)$ . The (log-)likelihood function then becomes

$$L(c_1, c_2 | R) = \prod_{i=0}^n e^{-\theta_1 c_1 - \theta_2 c_2} \left(\frac{\theta_1 c_1}{\theta_2 c_2}\right)^{r_i/2} I_{r_i}\left(2\sqrt{\theta_1 c_1 \theta_2 c_2}\right), \quad (4.2)$$

$$l(c_1, c_2 | R) = \sum_{i=0}^n -\theta_1 c_1 - \theta_2 c_2 + \frac{r_i}{2} \log\left(\frac{\theta_1 c_1}{\theta_2 c_2}\right) + \log\left(I_{r_i}\left(2\sqrt{\theta_1 c_1 \theta_2 c_2}\right)\right). \quad (4.3)$$

We can now take the derivative with respect to  $c_1$  and  $c_2$ . Before we do so, we need to know what the derivative of the modified Bessel function of the first kind is. That is given in the following proposition, see [2].

**Proposition 4.4.1.** The derivative of the modified Bessel function of the first kind  $I_y(\theta)$  can be written as

$$\frac{\partial I_y(\theta)}{\partial \theta} = \frac{y}{\theta} I_y(\theta) + I_{y+1}(\theta).$$

We can now take the derivative of (4.3) with use of proposition 4.4.1. We get

$$\frac{\partial l(c_1, c_2|R)}{\partial c_1} = \sum_{i=0}^n -\theta_1 + \frac{r_i}{2c_1} + \frac{\theta_1\theta_2c_2}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{\frac{r_i}{2\sqrt{\theta_1c_1\theta_2c_2}}I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2}) + I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right), \quad (4.4)$$

$$= \sum_{i=0}^n -\theta_1 + \frac{r_i}{2c_1} + \frac{r_i}{2c_1} + \frac{\theta_1\theta_2c_2}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right), \quad (4.5)$$

$$= \sum_{i=0}^n -\theta_1 + \frac{r_i}{c_1} + \frac{\theta_1\theta_2c_2}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right), \quad (4.6)$$

and

$$\frac{\partial l(c_1, c_2|R)}{\partial c_2} = \sum_{i=0}^n -\theta_2 - \frac{r_i}{2c_1} + \frac{\theta_1\theta_2c_1}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{\frac{r_i}{2\sqrt{\theta_1c_1\theta_2c_2}}I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2}) + I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right), \quad (4.7)$$

$$= \sum_{i=0}^n -\theta_2 + \frac{r_i}{2c_2} - \frac{r_i}{2c_2} + \frac{\theta_1\theta_2c_1}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right), \quad (4.8)$$

$$= \sum_{i=0}^n -\theta_2 + \frac{\theta_1\theta_2c_1}{\sqrt{\theta_1c_1\theta_2c_2}} \left( \frac{I_{r_i+1}(2\sqrt{\theta_1c_1\theta_2c_2})}{I_{r_i}(2\sqrt{\theta_1c_1\theta_2c_2})} \right). \quad (4.9)$$

Now we set (4.6) and (4.9) equal to zero. If we multiply (4.6) by  $c_1$  and (4.9) by  $c_2$  and then subtract them from each other, we get

$$\sum_{i=0}^n -\hat{c}_1\theta_1 + r_i + \theta_2\hat{c}_2 = 0,$$

and due to the independence of  $\hat{c}_1$  and  $i$ , we could express  $\hat{c}_1$  as

$$\hat{c}_1 = \frac{\sum_{i=0}^n r_i + \theta_2\hat{c}_2}{\sum_{i=0}^n \theta_1}. \quad (4.10)$$

Now we substitute (4.10) in (4.9) and set the equation equal to zero. We then get a equation with one unknown variable that can be solved. Solving this equation gives a value of  $\hat{c}_2$  which can be used to get  $\hat{c}_1$  with the use of (4.10).

### 4.3 Fisher information of Poisson difference distribution

We again want to know what the accuracy of our estimated  $\hat{c}_1$  is. This can again be done with a confidence interval. Since we have a probability mass function that is dependent on two parameters ( $c_1$  and  $c_2$ ), we must use the Fisher matrix as described in section 3.6. The Fisher information matrix can be calculated by making use of (3.9). To keep things short, we define  $\xi = 2\sqrt{\theta_1c_1\theta_2c_2}$ ,



where  $\theta_1$  and  $\theta_2$  have the same definition as in the previous section. Using straightforward derivative rules in combination with proposition 4.4.1 gives us

$$\begin{aligned}\mathcal{I}_{11}(c_1, c_2) &= \sum_{i=0}^n \frac{r_i}{c_1} + \frac{1}{2} c_1^{-\frac{3}{2}} \sqrt{\theta_1 \theta_2 c_2} \left( \frac{I_{r_i+1}(\xi)}{I_{r_i}(\xi)} \right) + \frac{\theta_1 \theta_2 c_2}{c_1} \left( \frac{\xi^{-1} I_{r_i+1}(\xi) I_{r_i}(\xi) + I_{r_i+2}(\xi) I_{r_i}(\xi) - I_{r_i}^2(\xi)}{I_{r_i}^2(\xi)} \right), \\ \mathcal{I}_{22}(c_1, c_2) &= \sum_{i=0}^n \frac{1}{2} c_2^{-\frac{3}{2}} \sqrt{\theta_1 \theta_2 c_1} \left( \frac{I_{r_i+1}(\xi)}{I_{r_i}(\xi)} \right) + \frac{\theta_1 \theta_2 c_1}{c_2} \left( \frac{\xi^{-1} I_{r_i+1}(\xi) I_{r_i}(\xi) + I_{r_i+2}(\xi) I_{r_i}(\xi) - I_{r_i}^2(\xi)}{I_{r_i}^2(\xi)} \right), \\ \mathcal{I}_{12}(c_1, c_2) &= \sum_{i=0}^n -\frac{1}{2} \sqrt{\theta_1 \theta_2 c_1 c_2} \left( \frac{I_{r_i+1}(\xi)}{I_{r_i}(\xi)} \right) + \theta_1 \theta_2 \left( \frac{\xi^{-1} I_{r_i+1}(\xi) I_{r_i}(\xi) + I_{r_i+2}(\xi) I_{r_i}(\xi) - I_{r_i}^2(\xi)}{I_{r_i}^2(\xi)} \right), \\ \mathcal{I}_{21}(c_1, c_2) &= \mathcal{I}_{12}(c_1, c_2).\end{aligned}$$

By making use of the Fisher information matrix and theorem 3.6, we can make a confidence interval.

## 4.4 Confidence interval

We return to our original problem, wherein we wanted to determine a confidence interval for  $c_1$ . We will use the same model as described in section 3.3. If we apply our method that was described above to six of those data sets, we get the following 95% confidence intervals for  $c_1$ ,

Data set	Lower bound	MLE	Upper bound	width
1	1.76	2.02	2.29	0.53
2	1.89	2.11	2.33	0.44
3	1.86	2.14	2.41	0.55
4	1.95	2.17	2.39	0.44
5	1.60	1.87	2.13	0.53
6	1.76	2.00	2.24	0.48

Table 4.1: 95% confidence intervals for  $c_1$ . All values should be multiplied by  $10^{-3}$ .

Now we can make a comparison with table 3.1. Both tables represent 95% confidence intervals for  $c_1$  with data sets from the same model. We can clearly see that the estimation from section 3.3 a better estimation gives. This is due to the fact that the width in table 3.1 is smaller than in 4.1, which is also what we would expect. The method from section 3.3 used the number of reactions that occurred, since we assume this is a known value. In the second method we assumed that we can only measure the change in concentration and since (2.1) and (2.2) are exact opposites, we were not able to determine the exact number of reactions. Therefore, the last method had less information to work with and that is why we would expect that the method from section 3.3 would give a more accurate estimation of  $c_1$  than with the method that uses the Poisson difference distribution.

## Chapter 5

# Aggregation

In this chapter we will take a look at a certain type of process, which is known as aggregation. The definition of the aggregation model is based on the book ‘A kinetic View of Statistical Physics’ [11].

In the aggregation process clusters merge with each other irreversibly. This phenomenon is observed in many different processes, i.e. cheese making and star formation. If we have  $n$  types of clusters, then the aggregation model consist of the reactions



with  $i \leq j$  and  $i + j \leq n$ .

To illustrate this model, we look at some examples. The first example we look at is the example of soap bubbles. Soap bubbles can merge with each other and then become a bigger bubble until they eventually burst. This fits the aggregation model, since clusters, in this case bubbles, can react with each other and form bigger clusters which are also bubbles. Another example is particle aggregation, where you start with many individual particles that are ‘sticky’. If two particles collide, they stick together and become a larger particle. The mass of the product will equal the sum of the mass of the reactants. It is common in aggregation that  $S_1$  is in the starting situation the only cluster which has a positive concentration or quantity.

In aggregation, the maximum size of the cluster is very important. In some aggregation processes the maximum cluster size is limited. After that size the clusters can break down to smaller pieces. However, in particle aggregation, the size of the highest possible cluster can be as large as the sum of the clusters of all the particles. The following assumptions about the maximum cluster size can be made:

- (i) The maximum cluster size equals the sum of the cluster sizes in total.
- (ii) When the cluster becomes greater than size  $n$ , it breaks down in cluster of smaller sizes.
- (iii) When the cluster becomes greater than size  $n$ , it vanishes.

For the aggregation model we can also use the tau-leaping algorithm. The question that arises is: ‘What does  $c_{i,j}$  look like?’. In this case, the reaction speed depends on the sizes of the clusters.

The larger the size, the higher the chance that another cluster will collide with this cluster. It can be argued that the reaction speed is exponentially dependent on the sizes of the clusters.

We therefore assume that  $c_{i,j} = C \cdot (i \cdot j)^\alpha$  with  $C, \alpha > 0$  and  $i \neq j$ . If  $i = j$ , then we get  $c_{i,i} = \frac{C}{2} \cdot (i \cdot i)^\alpha$ . The last case can be explained in the following way. If  $X_i$  is the number of clusters of size  $i$ , we get that the number of pairs of  $(S_i, S_i)$  is equal to  $\frac{1}{2}X_i(X_i - 1)$ , which is approximately equal to  $\frac{1}{2}X_i^2$  for large enough  $X_i$ . The factor  $\frac{1}{2}$  thus corrects for the double counting of pairs that would otherwise occur. We conclude from the definition of  $c_{i,j}$  that clusters with a larger size are more reactive. We can conclude this because  $\alpha > 0$ .

If we want to apply the tau-leaping algorithm to the aggregation model, we need to choose  $\tau$  to be very small, since big clusters can be very reactive. Furthermore, it is important to pay attention to the assumption about the largest cluster size. We look at an example, to show how the aggregation model works and how the assumption (i) and (iii) can change the model drastically. The values can be found in the table below.

$C$	$\alpha$	$\tau$	$X_1(0)$
$1 \cdot 10^{-4}$	0.9	0.02	10000

Table 5.1: Initial values for an aggregation model

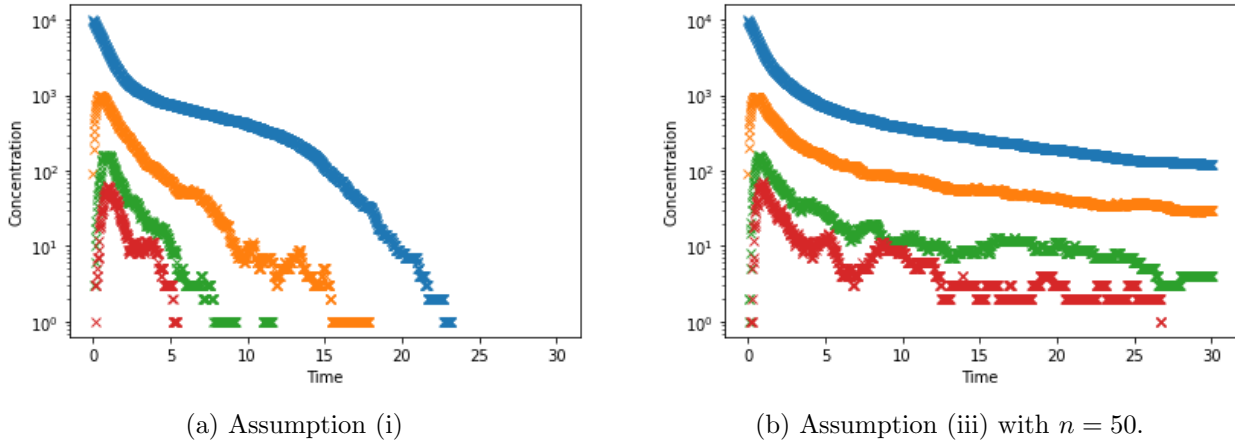


Figure 5.1: Implementation of the tau-leaping algorithm on the aggregation model,  $S_1$ : blue,  $S_2$ : orange,  $S_4$ : green,  $S_6$ : red.

We can see that  $S_1$  is decreasing in concentration, which is what we would expect, since  $S_1$  is never a product of a reaction. We also see that the concentrations of clusters of larger size are lower than the concentration of  $S_1$ . The main reason for this is that they are more reactive, because  $c_{i,j}$  is proportional to  $(ij)^\alpha$ . Another reason is, that there can exist less clusters of larger size than of smaller size, i.e. in the beginning the concentration of  $S_1$  is 10000, but that means that there can only exist a maximum of 10 clusters of size  $S_{1000}$ .

Another thing that stands out, is that the concentrations of  $S_1, S_2, S_4$  and  $S_6$  decrease much slower than with assumption (iii). This can also be explained. At the beginning, the graphs show the same behaviour. After  $t = 5$  they begin to differ drastically. The reason for this is that with assumption (i) big reactive clusters can exist, which will react with the clusters  $S_1, S_2, S_4$  and  $S_6$ , hence those clusters will decrease in concentration. It is therefore very important to really think through which assumption is suitable for a model, since the outcomes can be totally different.

## 5.1 Confidence interval parameters

Just like in the case of the previous chapters, we are again interested in estimating the parameters of the model, in this case  $C$  and  $\alpha$ . We suppose that the tau-leaping algorithm is a suitable fit and that the number of occurred reactions is known.<sup>1</sup> With those two assumptions, we can define the likelihood function as

$$L(C, \alpha | X) = \prod_{\mathcal{K}} \prod_{k=0}^n \frac{(X_i(\tau k) X_j(\tau k) \tau C \delta(i, j) (ij)^\alpha)^{r_k^{ij}}}{r_k^{ij}} e^{-X_i(\tau k) X_j(\tau k) \tau C \delta(i, j) (ij)^\alpha}, \quad (5.2)$$

with  $\mathcal{K} = \{(i, j) | 1 \leq i \leq n, i \leq j \leq n\}$ ,  $\delta(i, j) = \frac{1}{2}$  if  $i = j$  and  $\delta(i, j) = 1$  otherwise and  $r_k^{ij}$  the number of occurred reactions of type  $S_i + S_j \rightarrow S_{i+j}$  at time step  $k$ .

If we take the logarithm of (5.2) and then the derivative with respect to  $C$  and  $\alpha$  and let it equal 0, we get

$$\frac{\partial l(C, \alpha | X)}{\partial C} = \sum_{\mathcal{K}} \sum_{k=1}^n \frac{r_k^{ij}}{C} - (ij)^\alpha X_i(\tau k) X_j(\tau k) \tau \delta(i, j) = 0, \quad (5.3)$$

$$\frac{\partial l(C, \alpha | X)}{\partial \alpha} = \sum_{\mathcal{K}} \sum_{k=1}^n r_k^{ij} \log(ij) - C l \log(ij) (ij)^\alpha X_i(\tau k) X_j(\tau k) \tau \delta(i, j) = 0. \quad (5.4)$$

If we rewrite (5.3), we find

$$\hat{C} = \frac{\sum_{\mathcal{K}} \sum_{k=1}^n r_k^{ij}}{\sum_{\mathcal{K}} \sum_{k=1}^n (ij)^{\hat{\alpha}} X_i(\tau k) X_j(\tau k) \tau \delta(i, j)}. \quad (5.5)$$

If we substitute (5.5) into (5.4), we find

$$\sum_{\mathcal{K}} \log(ij) \sum_{k=1}^n \left( r_k^{ij} - \frac{\sum_{\mathcal{K}} \sum_{k=1}^n r_k^{ij}}{\sum_{\mathcal{K}} \sum_{k=1}^n (ij)^{\hat{\alpha}} X_i(\tau k) X_j(\tau k) \tau \delta(i, j)} (ij)^{\hat{\alpha}} X_i(\tau k) X_j(\tau k) \tau \delta(i, j) \right) = 0, \quad (5.6)$$

which is an equation that does not depend on  $\hat{C}$  anymore and has only one unknown variable, namely  $\hat{\alpha}$ . When we solved the equation, we can use the the value of  $\hat{\alpha}$ , for determining the value

<sup>1</sup>If we assume that we can only measure the change in concentration, then the method from chapter 4 has to be applied. This gives, however, some unpleasant expressions.

of  $\widehat{C}$  by using (5.5).

To calculate the confidence interval for  $C$  and  $\alpha$ , we need to use the Fisher information matrix again, since we have two parameters. Using (3.9) gives us

$$\mathcal{I}_{11}(C, \alpha) = \sum_{\mathcal{K}} \sum_{k=1}^n \frac{r_k^{ij}}{C^2}, \quad (5.7)$$

$$\mathcal{I}_{22}(C, \alpha) = \sum_{\mathcal{K}} \sum_{k=1}^n C \log^2(ij) (ij)^\alpha X_i(\tau k) X_j(\tau k) \tau \delta(i, j), \quad (5.8)$$

$$\mathcal{I}_{12}(C, \alpha) = \sum_{\mathcal{K}} \sum_{k=1}^n \log(ij) (ij)^\alpha X_i(\tau k) X_j(\tau k) \tau \delta(i, j), \quad (5.9)$$

$$\mathcal{I}_{12}(C, \alpha) = \mathcal{I}_{21}(C, \alpha). \quad (5.10)$$

## 5.2 Implementation

The method as described above can be again implemented in Python. The hardest problem is solving (5.6), since it cannot be solved exactly. We will solve this equation by making use of the bisection method. The data sets that we will be using are from an aggregation model where the clusters disappear after size  $n = 50$ . This is pure practical, since calculating a value of (5.6) is in  $\mathcal{O}(n^2)$ , because of the first summation.<sup>2</sup>

The parameters that are used can be found in table 5.2. We will determine a 95% confidence interval for  $C$  and  $\alpha$  with the use of theorem 3.6.

$C$	$\alpha$	$\tau$	$X_1(0)$	$n$
$2 \cdot 10^{-4}$	0.95	0.02	10000	50

Table 5.2: Initial values for the aggregation model.

We can see from table 5.3 and table 5.4, that we can determine the values of  $C$  and  $\alpha$  for this example with a relatively high accuracy. This is mainly due to the amount of information that the data sets contain. There are many reactions happening in this example of aggregation, which means that there is also a lot of information that can be deducted from the data. The more information there is, the more accurate the estimation of the parameters will be.

---

<sup>2</sup>Most of the times, this is not really a problem for aggregation models. With use of smart data structures it is usually still doable to solve (5.6) for large  $n$ .

Data set	Lower bound	MLE	Upper bound	width
1	1.901	1.956	2.010	0.110
2	1.949	2.005	2.061	0.112
3	1.969	2.025	2.081	0.112
4	1.935	1.990	2.045	0.110
5	1.977	2.034	2.090	0.113
6	1.926	1.981	2.036	0.110

Table 5.3: 95% confidence intervals for  $C$ . All values should be multiplied by  $10^{-4}$ .

Data set	Lower bound $\alpha$	MLE $\alpha$	Upper bound $\alpha$	width
1	0.949	0.960	0.971	0.22
2	0.937	0.948	0.960	0.23
3	0.935	0.946	0.958	0.23
4	0.936	0.947	0.959	0.23
5	0.931	0.943	0.955	0.24
6	0.942	0.954	0.965	0.23

Table 5.4: 95% confidence intervals for  $\alpha$ .

## Chapter 6

# Percolation

In the previous chapters we have seen how we could estimate parameters using a certain data set. These data sets were generated with Python code that uses the tau-leaping algorithm. Processes in real life are clearly not just data sets from Python. Therefore, it is time to test our method of estimating parameters on real data. The data that we will be using is from a two dimensional percolation process on a geometric graph. Percolation processes are well studied and a lot is already known about them.

The percolation process that we will use works as follows. We start with a graph that only contains nodes. We then start with randomly adding edges. If nodes are connected with each other by an edge, then they form a cluster, which size is equal to the number of nodes that it contains.

It can be hypothesised that the percolation process fits the aggregation model. Larger clusters have more neighbours to which they can connect, hence they are more reactive just as in the aggregation model. The value of  $\alpha$  is highly dependable on the number of neighbours. This can be seen the following way. Assuming we have a graph that contains nodes that are all on one straight line, such that they can only have an edge with their neighbours. This means that every node has two neighbours. For big clusters the same number holds, since they can only have one neighbour on the left and one on the right. Therefore, we can see that big clusters are not more reactive than smaller ones. In other words, we would expect that  $\alpha = 0$ . However, when the number of neighbours of each node is higher than two, then larger clusters do have, in general, more neighbours. In that case you would expect that  $\alpha > 0$ . The model stops when all nodes have formed one big cluster, just like in aggregation processes. How this takes shape can be seen in figure 6.1.

We assume that the aggregation model as described in chapter 5 fits the data from the two dimensional percolation. In this case the clusters can get as large as the sum of all the nodes at the beginning. The data that we will be using has been given by Dr. I. Kryven.

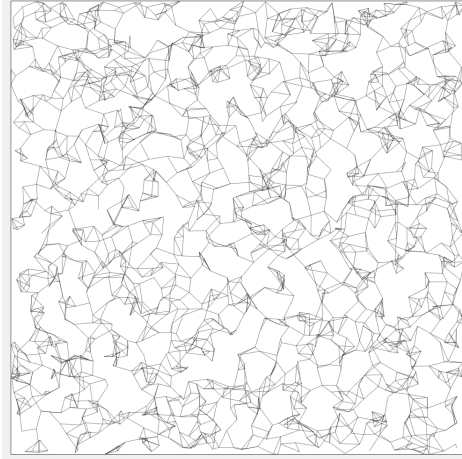


Figure 6.1: Final situation of percolation on a 2D geometric graph.

## 6.1 Confidence interval

The data that we will be using is from percolation on a two dimensional geometric graph with 2000 nodes. The data consists of 157 files at which is stated which clusters are formed at a specific time. We also have a file containing the information about the number of each cluster that exists at that time. The used time in the data set is scaled. The scaled time, let us call it  $\tilde{t}$ , is defined as the total number of nodes minus the number of clusters. Hence, the time steps are not uniform. This poses for us a problem, since we cannot directly implement the same method that we have seen in chapter 5, because that is based on fixed non-scaled time steps. However, there is a way we could work around this, but then we first have to make an assumption about the way that the data was measured. We assume that the reactions are counted every one time unit  $\tau$  and that then the  $\tilde{t}$  is calculated. What this means is that the time steps are not fixed given  $\tilde{t}$ , but are fixed given the normal non-scaled time steps. The steps are then equal to  $\tau$ . There is, however, still one problem. When in some time step  $\tau$  no reaction occurred, the time step will be 0 under  $\tilde{t}$ . This means that, when no reaction occurred, a time step in  $\tilde{t}$  can take longer than  $\tau$ , since the time  $\tilde{t}$  only gets updated when a reaction took place. Therefore we can never be sure, how long a time step in  $\tilde{t}$  takes in our standard non-scaled time.

From the data set we can see that there are several reactions occurring at each time step for  $\tilde{t} < 1600$ . We assume that for  $\tilde{t} < 1600$  in every time step  $\tau$  at least one reaction occurred. We now can state that every time step for which  $\tilde{t} < 1600$  takes  $\tau$  non-scaled time.

For the confidence interval of  $C$  and  $\alpha$  we can apply the method that is described in chapter 5. We will ignore everything that happens after  $\tilde{t} = 1600$ , since we cannot be certain what the lengths of the time steps are. The first step is to solve equation (5.6) and then we can determine  $\hat{C}$  by expression (5.5). Lastly we need to compute the Fisher information matrix with use of expressions (5.7)-(5.10). We find  $\hat{C} = 2.47 \cdot 10^{-5}$  and  $\hat{\alpha} = 0.482$ . The 95% confidence interval for  $C$  is  $[2.34 \cdot 10^{-5}, 2.60 \cdot 10^{-5}]$  and for  $\alpha$  we get  $[0.444, 0.519]$ .



## 6.2 Goodness of fit

The maximum likelihood estimators that we have calculated do not tell us anything about the goodness of fit of the aggregation model. We will determine if the aggregation model fits by comparing plots with each other. One plot is from the data itself, the other is an aggregation model that is modelled with the use of the tau-leaping algorithm and the maximum likelihood values of  $\alpha$  and  $C$ . First we will see how the concentrations of  $S_1, S_2, S_4$  and  $S_6$  change over time. Note that both plots have the scaled time on the x-axis.

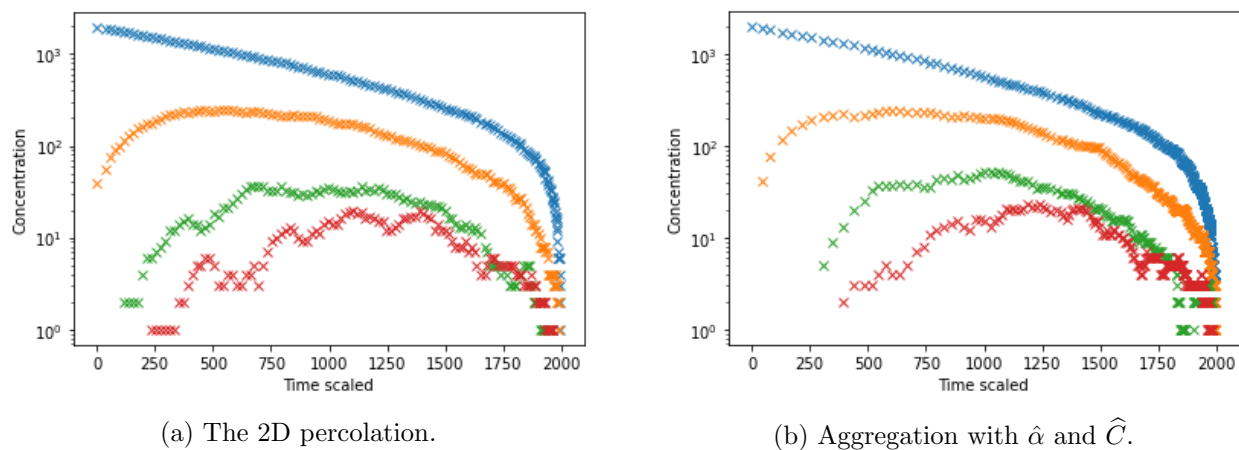
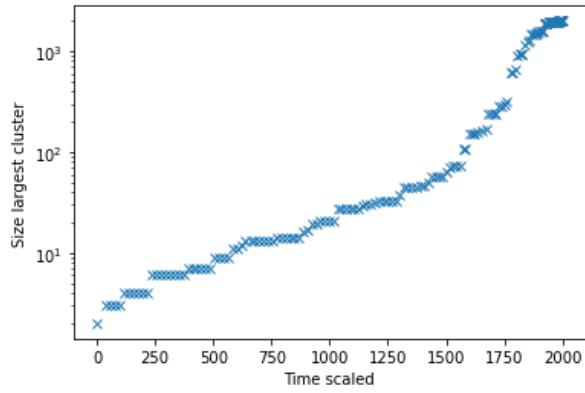


Figure 6.2: The two dimensional percolation and the aggregation process,  
 $S_1$ : blue,  $S_2$ : orange,  $S_4$ : green,  $S_6$ : red.

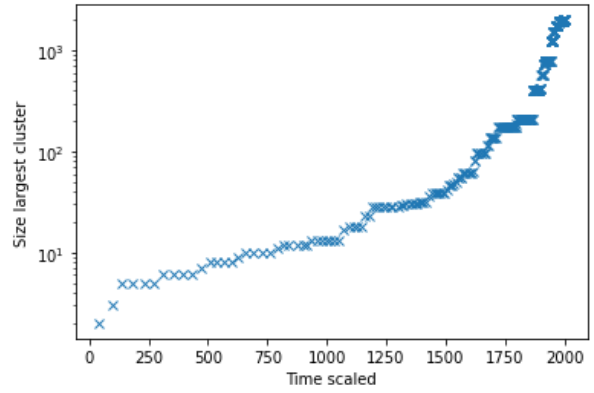
From figure 6.2a and 6.2b we can see two interesting things. First of all, we see that both figures show the same behaviour. This gives us an indication that the aggregation model fits the percolation process. Another thing that can clearly be observed is some interesting behaviour at the end of both plots. This is due to the time scaling. At the end of the process, one big cluster of size 2000 has formed, we know then that  $\tilde{t} = 2000 - 1 = 1999$ . Therefore you see a drastic decline in concentration in the end for  $S_1, S_2, S_4$  and  $S_6$ . This is only due to the scaled time, on a non-scaled time axis, you would not see this behaviour.

Lastly, we can also compare the largest cluster at each time.

Again we can see from figure 6.3a and 6.3b that both plots show the same behaviour, that indicates again that the aggregation model might be a good fit for the percolation. Another interesting thing that can be noticed, is that big clusters that has a size greater than 100 are only observable after  $\tilde{t} = 1500$ .



(a) The 2D percolation.



(b) Aggregation with  $\hat{\alpha}$  and  $\hat{C}$ .

Figure 6.3: The 2D percolation and a implementation of the tau-leaping algorithm on the aggregation model.

## Chapter 7

# Discussion

The methods that we have used in this paper strongly rely on the accuracy of the tau-leaping algorithm. Theoretically, the tau-leaping algorithm gives an exact result when the time steps are infinitesimally small. In real life however, the time steps that are taken to measure the concentration are not infinitesimally small. If the time steps are too big, then the tau-leaping algorithm does not give a good representation of the situation. If that is the case, then the methods of estimating parameters that are described in this paper are no longer reliable. Especially in aggregation we can have large deviations from reality. Large clusters in aggregation processes can be very reactive, but it is still important that the cluster for the most part does not react more than once in a single time step. This is only the case when  $\tau$  is sufficiently small.

If  $\tau$  is sufficiently small, then the tau-leaping algorithm is a good representation of the reality. We can then make some useful estimations. We can estimate the value  $\alpha$  for aggregation processes using the method from chapter 5. Usually with aggregation processes, the value of  $\alpha$  can be determined by reasoning, however now it can be checked if the reasoning is in line with the reality. It is important to note that there are several other ways for determining the maximum likelihood estimator, for example by the Cross-Entropy Method [12].

Estimating those parameters can be very useful in many ways. There is for example a strong relation between the tau-leaping algorithm and some ordinary differential equations that can describe the system at a macroscopically level. It turns out that the parameters that we estimated are related to the coefficients of the differential equations [1].

We have also seen that it is not necessary to know the number of reactions that occurred in order to estimate the parameters of a reaction process. By using the Poisson difference distribution it becomes possible to estimate reaction speeds by just measuring the change in concentration. This insight could lead to some new ideas in the field of transition theory.

Another thing that we encountered was the two dimensional percolation process. We have seen that there are some indications that the percolation process can be modelled with the use of an aggregation process. This was only based on similarities between some plots. To prove that the aggregation model fits the percolation model, more research has to be done which should have a better mathematical foundation.

## Chapter 8

# Conclusion

At the beginning of this paper we have seen how we could model certain kinds of reaction processes in which the reaction speed is dependent on the concentration of the reactants. As an illustration we have used an example with molecules, however, there are plenty of other processes where here could be used. The tau-leaping algorithm was one of the algorithms that we could use for modelling those kinds of processes. For small time steps, the algorithm gives us a trustworthy result. When we assumed that a process could indeed be modelled by the tau-leaping algorithm for sufficiently small time steps, then we could use the idea behind the algorithm to determine the maximum likelihood estimators and confidence intervals for the parameters.

From the confidence intervals that we constructed for some example processes, it became clear that, with use of the tau-leaping algorithm it is possible to estimate the parameters with quite some certainty. This is not generally the case and the certainty of the estimation is highly dependent on the data and the process itself. Finally, we have seen what aggregation processes are and how we could estimate the parameters of those processes. We also used the aggregation model on a two dimensional percolation process. We have seen that there are some indications that the aggregation model is indeed a good fit for the percolation process.

# Bibliography

- [1] Desmond J. Higham. “Modeling and Simulating Chemical Reactions”. In: *SIAM Review*. Vol 50, No 2, p. 347-368.
- [2] Abdulhamid A. Alzaid and Maha A. Omair. “On the Poisson difference distribution inference and applications”. *Bulletin of the Malaysian Mathematical Sciences Society*. Second Series 33.1 (2010), p. 17-45.
- [3] F.M. Dekking et al. “A Modern Introduction to Probability and Statistics”. Springer (2005), p. 313-322.
- [4] John A. Rice, “Mathematical Statistics and Data Analysis”. Cengage Learning (2006), chapter 8.
- [5] Andreas N. Philippou, and George G. Roussas. “Asymptotic normality of the maximum likelihood estimate of the independent not identically distributed case.” *Annals of the Institute of Statistical Mathematics* 27(1), p. 45-55 (1975).
- [6] D. Majerek et al. “Conditional strong law of large number”. *International Journal of Pure and Applied Mathematics*, Vol 20, No 2 (2005), p. 143-156.
- [7] A. Ben-Israel. “A Newton-Raphson method for the solution of systems of equations”. *Journal of Mathematical analysis and applications*, Vol 15, No 2 (1965), p. 94-98.
- [8] J. G. Skellam. “The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations”. In: *Journal of the Royal Statistical Society, Series A*, Volume 109 (1946), p. 296.
- [9] M. L. Esquível. “Probability generating functions for discrete real-valued random variables”. In: *Teor. Veroyatnost. i Primenen* . Volume 52, Issue 1 (2007), p. 129-149.
- [10] Milton Abramowitz and Irene A. Stegun. “Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables”. Vol 55, DC:National Bureau of Standards Applied Mathematics (1964), p. 376.
- [11] Pavel L. Krapivsky et al. “A Kinetic View of Statistical Physics”. Cambridge University Press (2010), chapter 5.
- [12] J. Revell, P. Zuliani. “Stochastic rate parameter inference using the cross-entropy method”. In: *International Conference on Computational Methods in Systems Biology*. Springer (2018), p. 146–164.