EXTENSION OF LEAP CONDITION IN APPROXIMATE STOCHASTIC SIMULATION ALGORITHMS OF BIOLOGICAL NETWORKS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SALİHA DEMİRBÜKEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
SCIENTIFIC COMPUTING

SEPTEMBER 2021

Approval of the thesis:

EXTENSION OF LEAP CONDITION IN APPROXIMATE STOCHASTIC SIMULATION ALGORITHMS OF BIOLOGICAL NETWORKS

submitted by **SALİHA DEMİRBÜKEN** in partial fulfillment of the requirements for the degree of **Master of Science in Scientific Computing Department, Middle East Technical University** by,

Prof. Dr. Sevtap Selçuk-Kestel _____
Director, Graduate School of **Applied Mathematics**

Assoc. Prof. Dr. Hamdullah Yücel _____
Head of Department, **Scientific Computing**

Prof. Dr. Vilda Purutçuoğlu _____
Supervisor, **Statistics, METU**

Prof. Dr. Ömür Uğur _____
Co-supervisor, **Scientific Computing, METU**

Examining Committee Members:

Prof. Dr. Tolga Can
Deparment of Computer Enginering, METU _____

Prof. Dr. Vilda Purutçuoğlu
Department of Statistics, METU _____

Prof. Dr. Ömür Uğur
Institute of Applied Mathematics, METU _____

Assist. Prof. Dr. Hajar Farnoudkia
Department of Business Administration, Başkent University _____

Dr. Melih Ağraz
Department of Applied Mathematics, Brown University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: SALİHA DEMİRBÜKEN

Signature :

ABSTRACT

EXTENSION OF LEAP CONDITION IN APPROXIMATE STOCHASTIC SIMULATION ALGORITHMS OF BIOLOGICAL NETWORKS

Demirbüken, Saliha

M.S., Department of Scientific Computing

Supervisor : Prof. Dr. Vilda Purutçuoğlu

Co-Supervisor : Prof. Dr. Ömür Uğur

September 2021, 61 pages

The approximate stochastic simulation (ASS) algorithms are the alternative approaches to generate the complex biological systems with a loss in accuracy by gaining from computational demand. There are a number of approximate methods which can successfully simulate the systems, such as poisson tau-leap and approximate Gillespie algorithms. The common property of these approaches is that they are based on the leap conditon which controls the change in hazard functions under a time interval. By means of this interval we can find an interval for the number of simultaneous reactions k in the time interval generated from the leap condition. In this study, we propose alternative intervals for k that are dependent on the sufficient statistics and also, we derive confidence intervals for k whose parameters are obtained via maximum likelihood estimator, moment estimators and bayesian estimators. Furthermore, we extend the leap condition by using higher order Taylor expansion whose original estimators are found under the first order. In our derivations, we use the poisson tau-leap approach since it is the fundamental approximate stochastic simulation method. Moreover, we apply the approximate Gillespie algorithm since it is one of the recent approaches that is derived from the idea of the poisson tau-leap method. We consider that although the proposal approaches to these two algortihms, they can be also adapted to other algorithms whose derivation are based on the parametric assumption.

Keywords: Approximate Stochastic Simulation Algorithms, Leap Condition, Confidence Interval

ÖZ

BİYOLOJİK AĞLARDAKİ YAKLAŞIK STOKASTİK SİMULASYON ALGORİTMALARINDA SIÇRAMA KOŞULUNUN GENİŞLETİLMESİ

Demirbüken, Saliha
Yüksek Lisans, Bilimsel Hesaplama Bölümü
Tez Yöneticisi : Prof. Dr. Vilda Purutçuoğlu
Ortak Tez Yöneticisi : Prof. Dr. Ömür Uğur

Eylül 2021, 61 sayfa

Yaklaşık stokastik simülasyon (YSS) algoritmaları , karmaşık biyolojik sistemleri yaratmak için hesaplama zamanı bakımından kazançlı olurken, doğruluklarından kaybeden alternatif yaklaşımlardır. Poisson tau-sıçraması ve yaklaşık Gillespie algoritması gibi sistemleri bu şekilde başarılı bir biçimde simüle edebilen bir dizi yaklaşık yöntem vardır. Bu yaklaşımların ortak özelliği, belirli bir zaman aralığında hazard fonksiyonlarındaki değişimi kontrol eden sıçrama koşuluna dayalı olmalarıdır. Bu aralık sayesinde, sıçrama koşulundan oluşturulan zaman aralığındaki k eşzamanlı reaksiyonlarının sayısı için bir aralık bulabiliriz. Bu çalışmada, yeterli istatistikte bağımlı olan k için alternatif aralıklar önerilmiş ve parametreleri maksimum olabilirlik tahmin edicisi, moment tahmin edicileri ve bayes tahmin edicileri ile elde edilen k için güven aralıkları türetilmiştir. Ayrıca, orijinal tahmin edicileri birinci dereceden bulunan daha yüksek dereceli Taylor açılımı kullanarak sıçrama koşulu genişletilmiştir. Türevlerimizde, temel yaklaşık stokastik simülasyon olduğu için poisson tau-leap yaklaşımını kullandık. Ayrıca, poisson tau-leap yöntemi fikrinden türetilen kalıcı yaklaşımlardan biri olduğu için yaklaşık Gillespie algoritmasını uyguladık. Önerinin bu iki algoritmaya yaklaşmasına rağmen, bunların türetilmesi parametrik varsayıma dayanan diğer algoritmalara da uyarlanabileceğini düşünüyoruz.

Anahtar Kelimeler: Yaklaşık Stokastik Simulasyon Algoritmaları, Sıçrama koşulu, Güven

Aralığı

To My Family

ACKNOWLEDGMENTS

I would like to deeply thank to all those people who supported me during my study. Firstly, I would like to express my very great appreciation to my thesis supervisor Prof. Dr. Vilda Purutçuoğlu for her patient guidance, enthusiastic encouragement and valuable advices during the development and preparation of this thesis. Despite of the Covid-19 Pandemic, her willingness to give her time and to share her experiences has brightened my path. I feel very lucky to be studying with her.

I also thank my co-advisor Prof. Dr. Ömür Uğur. He always answered kindly my questions. Moreover, he has contributed to the improvement of this study.

Additionally, I would like to thank the examining committe members, Prof.Dr. Tolga Can, Assist.Prof.Dr. Hajar Farnoudkia and Dr.Melih Ağraz for their valuable comments to advance my study.

I would also like to show my appreciation to my former supervisor Assoc. Prof. Dr. Hamdullah Yücel. I believe that he has contributed a lot to me, especially, in terms of both academic writing and mathematical perspective. Moreover, I am thankfull to everyone in the Institute of Applied Mathematics.

Furthermore, I want to thank to my friends Zeynep Sakartepe, Gizem Gürleyik, Aslıhan Kesekler for their friendship, support and help.

During the process of thesis writing, sometimes, I used the Caddepark Pharmacy. I would like to thank to everyone in the pharmacy due to their kindness and services.

Also, I thank to the METU research grant (Project No: 10282) for their support.

Lastly, I owe my most profound gratitude to my parents, my brother, his wife, my dear husband and his family for their continuous support through my entire life. Without your support, it is hard to study and accomplish. Especially, I would like to point out that my dear husband Mehmet Demirbüken. I appreciate him for his important help, motivation and love.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ASS	Approximate Stochastic Simulation
CME	Chemical Master Equation
MLE	Maximum Likelihood Estimator
MME	Method of Moment Estimator
ODE	Ordinary Differential Equation
SSA	Stochastic Simulation Algorithms

CHAPTER 1

PRELIMINARIES

1.1 Biological Modelling

The modelling makes people discover the elements of a system of interest, their states and their interactions with other elements. In detail, it is possible to model (some of) the mechanisms incluced transcription, translation, gene regulation, cellular signalling, DNA damage and repair processes, homeostating processes, the cell cycle, or apoptosis in the context of molecular cell biology. Moreover, it is possible to image the behaviour and time evolution of populations of individual organisms as well as the duty of a tissue, organ or even an entire organisms can be modelled [20].

In addition to these, to analyze the effect of interactions between the model components, biological modelling aims to collect a number of small models of well-understood mechanisms into a large model in data repository [20].

Why is stochastic modelling necessary? Although deterministic modelling can provide satisfactory informations about the biological systems, it cannot explain the randomless of the model. Therefore, we need to use more advance approach which is based on some advanced computational technology such as stochastical modelling [20].

Furthermore, when complex system are studied, the dynamics of the system are substantially stochastic, not deterministic [20]. As a basic example, let consider the linear birth-death process. The ordinary differential equation (ODE)

$$\frac{dY(t)}{dt} = (\lambda - \mu)Y(t) \tag{1.1}$$

stands for the change in the number of bacteria in the colony at time zero also known to be as y_0. Hereby, λ and μ denote the new individuals (birth) and the death rate per unit time, respectively. Then, the analytic solution of the given ODE in Equation (1.1) is found as $\lambda(t) = y_0 exp((\lambda - \mu)t)$. Here, there is a problem since the number of bacteria does not multiply continuosly and deterministicly. Actually, it is discretely and stochastically [20]. With more details, the level of molecular populations can alter considerably by discrete integer amounts, the time evolution of a biochemically reacting systems is continous [8].

Also, for the mathematical modelling, the system should be homogeneous. This is satisfied under the the conditions in which volume is constant, temperature and pressure are fixed [8]. Otherwise, it can be about diffusion process. In detail, if the reaction rate constant changes, the the temperature also changes. Because the reaction rate constant is based on the temperature [16].

1.2 Chemical and Biochemical Kinetics

A general form of the chemical reactions can be shown by the following way.

$$m_1\mathrm{A}_1 + m_2\mathrm{A}_2 + \ldots + m_r\mathrm{A}_r \longrightarrow n_1\mathrm{B}_1 + n_2\mathrm{B}_2 + \ldots + n_p\mathrm{B}_p, \tag{1.2}$$

where r is the number of reactants and p is the number of products. A_i refers to the ith reactant molecule and B_j presents the jth product molecule. Furthermore, m_i represents the number of molecules of A_i consumed in a single reaction step, while the subindex i is $i = 1, 2, \ldots, r$ and n_j indicates the number of B_j produced in a single reaction step for $j = 1, 2, \ldots, p$. Also, coefficients m_i and n_j are called as *stoichiometries* [20].

Basically, consider the system in below [9].

$$\begin{aligned} R_1 &: \mathrm{Y}_1 + \mathrm{Y}_2 \xrightarrow{\mathrm{k}_1} \mathrm{Y}_3 + \mathrm{Y}_4 \\ R_2 &: 2\,\mathrm{Y}_1 \xrightarrow{\mathrm{k}_2} \mathrm{Y}_1 + \mathrm{Y}_2 \\ R_3 &: \mathrm{Y}_1 \xrightarrow{\mathrm{k}_3} \mathrm{Y}_2, \end{aligned} \tag{1.3}$$

where R_j $(j = 1, 2, \ldots)$ means the reaction channel. In this expression, the quantity of each chemical is considered as a concentration. In detail, it is quantified in moles per litre , M, altering continuosly as the reaction continues. Finally, $[Y]$ is the notation for the concentration of a chemical species Y as shown in Equation (1.4) [20].

The ODEs for the system (1.3) can be shown by

$$\begin{aligned} \frac{d[Y_1]}{dt} &= -c_1[Y_1][Y_2] - c_2[Y_1] - c_3[Y_1] \\ \frac{d[Y_2]}{dt} &= -c_1[Y_1][Y_2] + c_2[Y_1] + c_3[Y_1] \\ \frac{d[Y_3]}{dt} &= c_1[Y_1][Y_2] \\ \frac{d[Y_4]}{dt} &= c_1[Y_1][Y_2], \end{aligned} \tag{1.4}$$

where c_1, c_2, c_3 are the rate constants which shows the speed of each reaction $(j = 1, 2, 3)$. Then,

$$v_1 = (-1, -1, 1, 1), v_2 = (-1, 1, 0, 0), v_3 = (-1, 1, 0, 0).$$

Here, the vector of v_{μ_i} denotes the change in the S_i molecular population caused by the occurence of one R_μ reaction [9]. Also, the system can be written as a matrix form as follows.

$$\frac{d}{dt}\begin{bmatrix}[Y_1]\\ [Y_2]\\ [Y_3]\\ [Y_4]\end{bmatrix} = \begin{bmatrix}-1 & -1 & -1\\ -1 & 1 & 0\\ 1 & 0 & 0\\ 1 & 0 & 0\end{bmatrix}_{4\times 3} \begin{bmatrix}c_1[Y_1][Y_2]\\ c_2[Y_1]\\ c_3[Y_2]\end{bmatrix}_{3\times 1}, \tag{1.5}$$

where the (4×3)- dimensional matrix is the stoichiometry matrix, V, of the reaction system [20]. The set of Equations (1.4), also called as the reaction rate equations (RREs), is the set of coupled, first order, ordinary differantial equations traditionally used to define the time evolution of a well-stirred chemically system [12].

Additionally, by setting the right hand side of the ODE (1.4) to be zero, it can be reached to the equilibrium solution of the system analyticly. In other words, the solution of the following system gives the equilibrium solution.

$$\begin{aligned} -c_1[Y_1][Y_2] - c_2[Y_1] - c_3[Y_2] &= 0,\\ -c_1[Y_1][Y_2] + c_2[Y_1] &= 0,\\ c_1[Y_1][Y_2] &= 0. \end{aligned}$$

Alongside being the reaction rate constants, k_i, in the deterministic systems, the stochastic system has stochastic rate constants c_i and associated rate law (hazard function), $h_i(y, c_i)$, where $Y = (Y_1, \ldots, Y_n)$ is the current state of the system. Also, $h_i(y, c_i)$ is determined by reaction R_i, respectively. Actually, $h_i(Y_i, \ldots, Y_N)$ gives the number of distinct combinations of R_i reactant molecules in the system when there are exactly Y_i molecules ($i = 1, \ldots, N$). For the system in (1.4) the following funtions can be written

$$\begin{aligned} h_1(Y, c_1) &= Y_1 Y_2\\ h_2(Y, c_2) &= \frac{Y_1(Y_1 - 1)}{2} \qquad (1.6)\\ h_3(Y, c_3) &= Y_1 \end{aligned}$$

(1.7)

Also, the propensity, i,e., the hazard, function $a_j(y)$ is found by $a_j(y) = c_j h_j(y)$. Therefore, it can be written the equality $a_j(y) = h_j(y, c_j)$. Then, $a_j(y)$ gives the the probability that an R_i reaction will happen in the time interval $(t, t + dt]$ [9], [10], [20]. That is,

$$\begin{aligned} a_j dt \triangleq &\text{ the probability, given } Y(t) = y, \text{ that are } R_j \text{ reaction will occur somewhere}\\ &\text{inside } \Omega \text{ in the next infinitesimal time interval } (t, t + dt]. \end{aligned} \tag{1.8}$$

From Equation (1.8), it can be obtained via the following probability.

$$P(y, t|y_0, t_0) \triangleq Prob\{Y(t) = y, \text{ given } Y(t_0) = y_0\} \tag{1.9}$$

By using Equation (1.8) and (1.9), the derivation below is acquired

$$\frac{\partial P(y, t|y_0, t_0)}{\partial t} = \sum_{j=1}^{M} [a_j(y - v_j)P(yx - v_j, t|y_0, t_0) - a_j(y)P(y, t|y_0, t_0)]. \tag{1.10}$$

In other words, it is a chemical master equation (CME) that indicates the probability that each species will get a specified molecular population at a given future [12].

Equation (1.10) can be solved occasionally for the probability density function of $Y(t)$. Here, as a novelty, a new probability function $p(\tau, j|y, t)$ is defined rather than using the function $P(Y, t|Y_0, t_0)$ to simulate the trajectories of $Y(t)$ versus t. The associated definition can be presented as below.

$p(\tau, j|y, t)d\tau \triangleq$ the probability, given $Y(t) = y$, that the next reaction in the system will occur in the infinitesimal time interval $[t + \tau, t + \tau + dt]$, and will be an R_j reaction.

Then,

$$p(\tau, j|y, t) = a_j(y)exp(-a_0(y)\tau), \tag{1.11}$$

where

$$a_0(y) \triangleq \sum_{j'=1}^{M} a_j(y). \tag{1.12}$$

Actually, Equation (1.11) is the mathematical basis for the stochastic simulation algorithm [12].

1.3 Leap Condition

The time step τ should be selected small enough that there is no significance change in the propensity function during the time interval $[t, t + \tau]$. In fact, the value of the function will not be altered by macroscopically noninfinitesimal. Actually, the propensity function will stay constant at the value of $a_j(Y)$ for each reaction channel R_j during the time interval $[t, t + \tau]$ for $j = 1, \ldots, r$. This implies that the probability that channel R_j will occur during any infinitesimal interval dt in $[t, t + \tau]$ is $a_j(Y)dt$. Then, in the history axis of the system, leaping down process can be in the following way by the amount of τ from state Y at time t. The number of times reaction channel R_j fires in $[t, t + \tau]$ is the value of k_j. A sample value of k_j can be generated from a random variable using some probability distribution such as poisson distribution or gamma distribution. Then, λ is the net change in the state of the system in $[t, t + \tau]$. λ can be described by $\lambda = \sum_{j=1}^{r} k_j v_j$, where v_j is the jth row of the stoichiometry matrix V. Thus, the leap condition satisfies that the following substraction $|a_j(Y + \lambda) - a_j(Y)|$ is effectively infinitesimal [11]. In other words, the leap condition can be shown by

$$|a_j(Y + \lambda) - a_j(Y)| \leq \epsilon a_0(Y).$$

Hereby, in this study, we aim to construct the confidence intervals for the population parameters k, which is the number of the reactions in the system, and the time step τ in the two

approximate SSAs, namely, the poisson tau-leap, which is the most recent approximate SSAs and its extension, which is called the approximate Gillespie method, by using maximum likelihood estimator (MLE), method of moment estimator(MME) and Bayesian approach. All these approaches are based on the leap condition. In the current literature, the k and τ in these simulation approaches have been used via a conservative one-sided confidence interval without controlling the significance level α. Hence, this study suggests realistic and accurate confidence intervals for both parameters by controlling α and by using the MLE, MME and Bayesian approach of the modal parameters so that narrower and more accurate confidence intervals can be obtained theoretically. Thereby, in the organization of this thesis, we introduce exact and approximate stochastic simulation algorithms in Chapter 2. We present our confidence intervals inserted to the leap condition of the underlying two approximate SSAs in Chapter 3. Finally, we conclude our findings and we present the our future work in Chapter 4.

CHAPTER 2

EXACT AND APPROXIMATED STOCHASTIC SIMULATION ALGORITHMS

In this study, we use poisson distribution and gamma distribution. In this view, although we focus on poisson τ-leap method and approximate Gillespie algotrithm, we want to present main exact and approximate stochastic simulation algorithms in the literature. These approaches are given in the following.

2.1 Stochastic Simulation Algorithms

Due to the Monte Carlo approach, it is possible to estimate the probability of varying outcomes if there is an interference of random variables. In other words, this method is used for the case where the outputs of the experiments are not sure [15].Stochastic simulation algorithm (SSA) is a Monte Carlo procedure to calculate numerically time trajectories of the molecular populations in exact accordance with the chemical master equation (CME) [13]. However, the SSA is very slow since it forces to simulate every different reaction event [12]. Moreover, it is just available for distinctly finite chemical systems. In other words, to simulate the system, there should be limited on the total number of molecules in the system. For some cases, this sides can become an advantage rather than disadvantage because of having a local connection in the chemical reactions for the systems with huge numbers of population. Furthermore, SSA gives exact solution because unlike numerical approaches, infinitesimal time increment dt is never approximated by using finite time step Δt [9]. Below the major SSAs, namely, the direct method, first reaction method and next reaction method are explained in detail in the following part.

2.1.1 Direct Method (Gillespie Algorithm)

This method is also called as *the Gillespie Algorithm*. It investigates for a biological system answers via two questions [8].

1. Which reaction occurs next?

2. When does it occur?

Those questions can be answerable by identifying a reaction probability denstiy function on the space of continuous time random variable τ with $0 \leq \tau < \infty$ and discrete reaction indicator variable j with $j = 0, 1, \ldots, r$ showing a density function as

$$P(\tau, j) = \begin{cases} h_j(Y)exp(-h_0(Y)\tau) & ,0 \leq \tau < \infty \\ 0 & ,otherwise, \end{cases}$$

where $h_j(Y) = a_j c_j$ and $h_0(Y) = \sum_{j=1}^{r} h_j(Y) = \sum_{j=1}^{r} a_j c_j$, a_j and c_j are the number of the distinct molecular combinations of the given state Y and the stoichiometric reaction rate constant of the jth reaction, respectively [17]. In this expression, c_j has almost the same meaning with respect to the rate of constants in Equation (1.5) in the sense that it indicates the speed of the reaction.

By the Gillespie Algorithm, a random value is chosen from exponential distribution with rate $h_0(Y)$. In other words, time to the next reaction is $\tau \sim Exp(h_0(Y))$. Then, the reaction type j is chosen randomly with $\frac{h_j/Y}{h_0(Y)}$. Finally, the update of the system continue by using the time to next event and the event type.
In brief, the steps of the algorithm can be followed as [17],[20];

1. Initialization: Initialize at $t = 0$ with rate constants $c_1, \ldots, c_r$ and initial numbers of molecular populations for each species $Y_1, \ldots, Y_n$ for time t, where r and n describe the total number of reactions and substrates, i.e, species in the system, respectively.
2. Calculate $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$ by evaluating $h_j(Y) = a_j(Y)c_j$, which relies on current state Y for each $j = 1, 2, \ldots, r$.
3. Simulate the time to the next event, τ, developed from $Exp(h_0(Y))$ and the reaction index, $j = 1, 2, \ldots, r$, with probability $\frac{h_j(Y)}{h_0(Y)}$.
4. Install $t := t + \tau$, then update Y with respect to j. In other cases, update $Y := Y + Y_j$ describes the j^{th} column of the stoichiometry matrix V.
5. If $t < T_{max}$, where T_{max} is a total time period, return to step 2.

2.1.2 First Reaction Method

The difference between First Reaction method and the Direct method is that the latter generates the reaction indicates j and the continuous time τ directly, nonetheless, the former produces a presumed time τ_j for the reaction j occuring if no other reaction happened first [17]. Generally, since the first reaction method is computationally less efficient than the Gillespie Algorithm, the usage of this method is seldom when there are many reaction channels in

the system [11], [12]. In addition to these, like the direct method, in this method, the same probability distribution is used to generate j and τ. Consequently, to update the hazard function $h_j(Y) = a_j c_{ji}$ for $j = 1, \ldots, r$ and to find the smallest τ_j in each iteration, r number of the variable, τ_j is generated by this method [17]. In summary, the algorithms for the first reaction method can be listed as below;

1. Initialization: Initialize the number of the molecules of each species and set $t = 0$.
2. Calculate the hazard function $h_j(Y) = a_j(Y)c_j$ for all j.
3. Generate (a presumed time) τ_j for each j with regard to an exponential distribution with parameter $h_j(Y)$, $\tau_j \sim Exp(h_j(Y))$.
4. Choose the least time step τ_j for the next reaction where j be the index of the smallest of $\{\tau_1, \ldots, \tau_r\}$.
5. Change the number of molecules considering j and τ.
6. Update the time t by $:= t + \tau$.
7. If $t < T_{max}$, go to step 2.

2.1.3 Next Reaction Method (Gibson-Bruck Algorithm)

This method is also known as "Gibson-Bruck Algorithm". Primarily, a change form of the first reaction method, the first reaction method is faster and more efficient than the Direct method. But it is more compelling to do. [12]. The basic consept of the method can be summarized as the following way. Instead of just storing hazard function $h_j(Y)$, both time step τ_j and $h_j(Y)$ are stored with sensitively recalculating $h_j(Y)$(and τ_j) if it changes. Moreover, investigating the set of reactions beforehand and deciding which reactions after which $h_j(Y)$ can be done by a dependency graph. This graph enables to renew the minimum number of $h_j(Y)$'s. Then, τ_i's is appropriately reused. Although, assuming statistically independent random numbers in the Monte Carlo simulations are generally not valid, here, it is rightful. Particularly, the study of Gibson and Bruck (200) [7] presents the proof of this case. Moreover, it can be possible to reuse all τ_i's except for τ_{ν_i}, which is the smallest putative time. On the other hand, the difference from the First Reaction method is to use absolutive times rather than to use the relative times. Thereafter, in order to update $h_j(Y)$ (and τ_j), by using proper data structure for saving them, an operation indexed priority queves is defined. [7].

The algorithm can be shown below [7],[17].

1. Initialization:
    a. Initialize both the numbers of molecules Y and rate of constant c.
    b. Calculate hazard functions $h_j(Y)$ for $j = 1, \ldots, r$.

c. Store the value of τ_j calculated, where exponential distribution with parameter $h_j(Y)$, i.e, $\tau_j \sim Exp(h_j(Y))$.

2. Let k be the index of the smallest τ_j, i.e., $k = \underset{j}{min}\{\tau_j\}$ and set $t = \tau_k$.

3. Update Y to show execution of the reaction k.

4. Update the $h_k(Y)$ with respect to new state Y and compute $\tau_k = t + Exp(h_k(Y))$, where $Exp(h_k(Y))$ shows the sample from the exponential distribution with rate $h_k(Y)$.

5. For each reaction j ($j \neq k$) whose hazard is changed by reaction k,

   a. Update $h'_j(Y) = h_j(Y)$ and keep the old $h_j(Y)$ provisionally.

   b. Calculate $\tau_j = t + \frac{h_j(Y)}{h'_j(Y)}(\tau_j - t)$.

   c. Delete the old $h_j(Y)$.

6. Repeat the step 4 for each reaction j ($j \neq k$).

7. Go to step 2 if $t < T_{max}$.

For further discussiun at this point, we can refer the study of Gibson and Bruck [7].

2.2 Approximate Stochastic Simulation Algorithms

Even though advancement of the stochastic simulation algortihm (SSA) identified previously gives helpful results, considering being efficient, the time for any process simulating every reaction event G is too slow. To speed the time of the simulation, there can be some sacrifice in the exactness of the SSA. Using approximate SSAs is one way to do this [12]. These fast algorithms depend on a time discretisation of the Markov process [20]. Here, the main concept is that, firstly, the time is seperated into small different pieces, called leap. Then, to be possible to continue with improvement of the state from the start of one piece to another, the basis kinetics are approximated [20]. Mostly, these algorithms act with the assumption of the leap condition. In other words, the selected time interval τ should be satisfied that there is no notable change in the propensity function during the time change from t to $t + \tau$. In addition to these, since SSAs are computational costly, approximate SSAs make possible to obtain computationally less demanding.

Below, we present the major approximate SSA and the idea of the leap condition which is the fundamental approach in the generation of the approximate approaches.

2.2.1 Poisson τ- Leap Method

Under the leap condition, the aim of this method is to enhance intervals between selecting times with chosen the time interval τ as large as possible [11], [17]. Here, for each reaction

channel R_j, a random value k_j is produced from a Poisson distribution by $Poi(h_j(Y)\tau)$ in the time interval $[t, t+\tau]$, where $Y(t) = Y$ is a state vector. Then, an acceptable τ is reached via substituting boundary for the difference between

$$|h_j(Y + \lambda(Y)) - h_j(Y)|,$$

where $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$ denotes the net change in the state of the system in $[t, t+\tau]$. As $k_j \sim Poi(h_j(Y)\tau)$, the mean of k_j equals to $h_j\tau$, i.e., $E(k_j) = h_j(Y)\tau$ and

$$\bar{\lambda}(Y, \tau) = \sum_{j=1}^{r} [h_j(Y).\tau] v_j = \tau \xi(Y) \tag{2.1}$$

that gives the expected net change in the state for the given time interval. In this equality, j represents the stoichiometric coefficients of the reaction j corresponding to the jth row of the net effect matrix V as defined beforehand and $h_j(Y)$ corresponds to the hazard function of the jth reaction which is found by the product of the rate constant c_j and distinct molecular reactant combination of underlying reaction. Subsequently, $\xi(Y) = \sum_{j=1}^{r} h_j v_j$ can be represented as the mean or expected state change in a unit of time by an n- dimensional vector where each ith component, $\xi_j(Y)$, subtends to the expected change of the ith species in an unit of time. Afterwards, the following inequality is obtained

$$|h_j(Y + \bar{\lambda})(Y, \tau) - h_j(Y)| \leq \epsilon h_0(Y) \tag{2.2}$$

by using $\lambda(Y, \tau)$ in Equation (2.1). It can be inferred from Equation (2.2) in the following way. The expected changes in the hazard functions in time τ is restricted by a fraction ϵ, error control parameter lying $0 < \epsilon < 1$, and the sum of all hazard functions $h_0 = \sum_{j=1}^{r} h_j(Y)$. In fact, this inequality presents a leap condition. Consequently, using the first order Taylor series expansion helps to predict the difference on the left hand side of Equation (2.2). After the application of the Taylor series expansion, the equality below is obtained.

$$|h_j(Y + \bar{\lambda}(Y, \tau)) - h_j(Y)| \approx \bar{\lambda}(Y, \tau) h_j(Y) = \sum_{i=1}^{n} \tau \xi_j(Y) \frac{\partial h_j(Y)}{\partial Y_i}. \tag{2.3}$$

Then, setting $b_{ji}(Y) = \frac{\partial h_j(Y)}{\partial Y_i}$ $(i = 1, \ldots, n; j = 1, \ldots, r)$ the following inequality can be shown

$$\tau | \sum_{i=1}^{n} \xi_j(Y) b_{ji}(Y) \leq \epsilon h_0(Y)|.$$

Consequently, the largest value of τ satisfying the leap condition for the given Y and the preselected ϵ is calculated by

$$\tau = min\left\{ \frac{\epsilon h_0(Y)}{|\sum_{i=1}^{n} \xi_j(Y) b_{ji}(Y)|} \right\}. \tag{2.4}$$

After all, using the exact SSA is more preferable rather than Equation (2.4) since the obtained value of τ is favorable for the leap size. The obtained value of τ in Equation (2.4) would not be efficient if $\tau \leq \frac{1}{h_0(Y)}$ despite the fact that it would be accurate. In fact, the time step is choosen by the approach same as SSA. Not considering of the computational cost, the time

interval in the Poisson τ-leap methos is more suitable than the time of SSA. Essentially, there is an incremental difference between them.

In final step, there is an update of the current state in the Poisson τ-leap method by replacing t by $t := t + \tau$. In addition, for Y, there is a necessity to determine the largest value of τ and to be adaptable with the leap condition.

Beside this, in the long-run simulations, Poisson τ-leap algorithm can create problem of negative molecular populations from the application of this method in various systems. To overcome this problem, there are some suggested solutions. One of the well-known alternative solution is the Binomial τ-leap approach. Although it does not give accurate result, to have smooth approximation of exact SSAs, it overcomes the negativity problem [14],[17],[5].

2.2.2 Langevin τ-leap Method

This approach states that if there is no significance change in the hazard functions $h_j(Y)'$s $(j = 1, \ldots, r)$, by the expected number of execution of reactions in the time step τ that satisfying the leap condition under the history axis of the system, then it is possible to write the Poisson distribution as a normal distribution with the same mean and variance when the number of molecules is large. That is, $k_j \sim \text{Poi}(h_j(Y)\tau) \approx N(h_j(Y)\tau, h_j(Y)\tau)$, where k_j is the number of times of execution of the jth reaction for $j = 1, \ldots, r$. So the steps of the Langevin τ-leap method can be presented as follows.

The first step is to select τ under the leap condition such that $\tau \gg max\{\frac{1}{h_j(Y)}\}$. The second step is to choose a sample value k_j by putting $k_j = h_j(Y)\tau + (h_j(Y)\tau)^{\frac{1}{2}} l_j$ where a sample value l_j is chosen from normal distribution with mean zero and unit variance, $l_j \sim N(0, 1)$, for each j, $(j = 1, \ldots, r)$. The following is calculate the net change in the state $\lambda(Y)$ by $\lambda(Y) = \sum_{j=1}^{r} k_j \nu_j$, where ν_j is the n-dimensional net effect vector of the jth reaction in n-species in the system. The last step is to replace $Y := Y + \lambda(Y)$ and $t := t + \tau$ and then to update the system.

In general, this method is successful to improve the speed of the Poisson τ-leap approach, it can be applicable if the system is large enough to assume the normal distribution in place of poisson density to describe the change in states. Also, since the leap condition is in the absolute value, the results cannot be negative values. Thus, it is preferable for very large systems with large number of per genes, per molecules or any species in the system and long-run simulation [17], [11].

2.2.3 Estimated Midpoint Techniques

It is neceassary that the chosen time step τ should not modify the propensity function under the leap condition. Nonetheless, it is possible to face with a problem of increasing compu-

tational errors when the chosen τ to have a faster simulation than the SSA may bring some changes the propensity functions. Consider the following example to see the similar problem in numerical calculations of the differential equation, while solving $\frac{dY}{dt} = f(y)$ by Euler method (a discretized version of the sde), where Δt is the alteration in tine along the t axis, $Y(t + \Delta t) = Y(t) + f(Y(t))\Delta t$ may generate computational errors whenever there is a change in f during the Δt. To minimize the errors, one way is to use the estimated-midpoint (or second order Runge-Kutta) algorithm. In this method, the first one is to calculate the change in Y as $\Delta_1 Y = f(Y(t))\Delta t$ and then to substitute it in the value of $Y(t + \Delta t)$, where $Y(t+\Delta t) = Y(t)+f(Y(t)+\frac{\Delta_1 Y}{2}Y)\Delta t$. After that, it is possible to calculate the actual increment in Y by regarding the slope function f at the estimated midpoint of Y during $[t, t + \Delta t]$ calculating from the Euler Method.
To get the expected state change $\bar{\lambda}(Y, \tau)$, it can be possible to customize this procesure to the Poisson τ-leap method. Hereby, $Y + \lceil \lambda(\bar{Y, \tau})/2 \rceil$ with $\lceil z \rceil$ denoting the largest integer in z corresponds the estimated-midpoint state during the leap τ.
There are two kinds of algorithms for the estimated midpoint tecniques [11],[17].

a. Estimated Midpoint τ-leap Method

Firstly, under the leap condition, calculate the expected state change $\lambda(\bar{Y}, \tau)$ by $\lambda(\bar{Y}, \tau) = \tau \sum_{j=1}^{r} h_j(Y)v_j$ during $[t, t + \tau]$. Hereby, $h_j(Y)$ and v_j are the hazard function and the n-dimensional stoichiometry vector of the jth reaction, respectively. Then, compute the expected new state $Y' = Y + [\lambda(\bar{Y, \tau})/2]$ and generate the number of execution time k_j from a Poisson distribution with mean $h_j(Y')\tau$ for each reaction $j = 1, \ldots, r$, i.e, $k_j \sim Poi(h_j(Y')\tau)$. Finally, obtain the net change in the state λ by $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$ and update the system by replacing Y and t with $Y + \lambda(Y)$ and $t + \tau$, respectively [11],[17].

b. Estimated Midpoint Langevin τ-leap Method

Under the leap condition and $\tau \gg max\{\frac{1}{h_j(Y)}\}$, it is possible to approximate Poisson distribution by normal distribution. Therefore, instead of the Estimated-midpoint τ-leap method in Section 2.2.3, the Estimated-midpoint Langevin τ-leap can be used. In this approach, under the given conditions, the first one is to compute the expected state change $\lambda(\bar{Y}, \tau) = \tau \sum_{j=1}^{r} h_j(Y)v_j$. Then, generate a sample value l_j coming from normal distribution with zero mean and unit variance, $l_j \sim N(0, 1)$, and calculate $k_j = h_j(Y')\tau + (h_j(Y')\tau)^{1/2} l_j$ for each $j = 1, \ldots, M$. The remaining steps of the algorithm are the same as the previous one. Particularly, compute the new expected state $Y' = Y + [\lambda(\bar{Y, \tau})/2]$ and the net change in the state $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$. Lastly, update the system $Y + \lambda(Y)$ and $t + \tau$, by replacing the state Y and the time t, respectively.

2.2.4 Binomial τ-leap Methods

Some τ-leap methods have a problem of generating negative populations. Particularly, it can be obtained negative populations in two ways. Firstly, the sample value for the reaction number could be greater than one of the molecular numbers in that reaction channel. Secondly, different reaction channel can occur simultaneously [19]. To overcome this problem, Binomial τ-leap approaches are suggested by Chatterjee and Vlachos (2005) [4] and Tian and Burrage (2004) [19]. According to Chatterjee and Vlachos (2005), conditions for the Binomial τ-leap method are the same as other τ-leap methods. In other words, there is a usage of the system's history axis to generate an approximated simulation under the leap condition. In the time interval $[t, t+\tau]$ given the state $Y(t)$, a maximum number of firings k^j_{max} of each reaction R_j is calculated by

$$k^j_{max} = min_{v_{ij}<0}\left\lceil \frac{Y_i}{|v_{ij}|} \right\rceil.$$

Hereby, Y shows the current state vector containing the number of molecules $Y_i(t)$ of all substrates ($i = 1, \ldots, n$) at time t, v_{ij} presents each entry of the net effect matrix for the ith species and the jth reaction ($j = 1, \ldots, r$) and also $\lceil z \rceil$ denotes the greatest integer in z. In addition, the hazard functions h'_js would be almost constant during $[t, t+\tau]$, where $h_j(Y) = h_j(Y(t))$ depends on $Y(t)$ [4]. To avoid the negative populations in Binomial τ-leap method, an upper bound is placed on the number of S_i molecules in the time interval $[t, t+\tau)$[4]. Under the leap condition, with k_j's from binomial distribution k^j_{max} maximum firings, each R_j firing has a success probability $p = \frac{h_j(Y)\tau}{k^j_{max}}$ (and a fail probability of $1-p$), where $\tau = \frac{f}{\sum_{j=1}^{r} h_j(Y)}$ and f indicates a coarse forming factor greater than 1. If $f = 1$, it stands for the average time increment of the exact SSA. In the method, it is possible to determine the time and the accuracy of the approximation by choosing $f < 10^3$ for small steps or $f > 10^4$ for large steps [17].

Moreover, according to Chatterjee and Vlachos(2005) [4], the probability of the number of firings k_j of R_j depending on the binomial distribution can be indicated mathematically in the following way.

$$P_{BD}(k_j; p, k^{(j)}_{max}) = \frac{k^{(j)}_{max}!}{k_j!(k^{(j)}_{max} - k_j)!} p^{k_j}(1-p)^{k^{(j)}_{max}-k_j}. \tag{2.5}$$

In brief, the steps of the algorithms of the Binomial τ-leap can be shown as follows [4].

1. Access the stoichiometric coefficients v_{ij}, initial population size $Y(0)$, and the rate constants c required in calculating the propensities.
2. Start the time at $t := t_0$.
3. Renew the steps $4-6$ up to obtain a maximum time, t_{max}.
4. Calculate the hazard function $h_j(Y(t))$ with $Y(t)$ and set $\tilde{Y} := Y$.

5. Choose admissable τ and update the time $t := t + \tau$.

6. For $j = 1$ to r reactions.

   a. Get $k_{max}^{(j)} = min_{v_{ij}<0}\lceil \frac{\tilde{Y}_i}{|v_{ij}|} \rceil$, where $\lceil z \rceil$ is the greater integer function.

   b. Sample k_j from Equation (2.5) as $p = \frac{h_j \tau}{k_{max}^j}$.

   c. Set $\tilde{Y}_i := \tilde{Y}_i + v_{ij} k_j$ for $i = 1, \ldots, n$ if $v_{ij} < 0$.

   d. Go to step 6a to compute $k_{max}^{(j)}$.

7. Update $Y_i := Y_i + v_{ij} k_j$ for $i = 1, \ldots, n$.

8. Return to step 3 .

Consequently, the method is a way to overcome the problem of negative populations in the unbounded Poisson τ-leap methods. The range of sample value is from zero to infitinity in the Poisson distribution. However, unlike the Poisson approaches, the range of sample value of $Bin(n, p)$ is an integer between 0 and n.

In addition to the study of Chatterjee and Vlachos (2005), Tian and Burrage (2004) suggested the Binomail τ-leap method to remove the problem of negative populations in the Poisson τ-leap approaches.

In this approach, the k_j is obtained from binomail distribution $Bin(N_j, \frac{h_j(Y)\tau}{N_j})$ under the condition

$$0 \leq \frac{h_j(Y)\tau}{N_j} \leq 1. \tag{2.6}$$

Hereby, N_j is defined as follows.

1. The first-order reaction
   $\mathrm{S}_1 \xrightarrow{\mathrm{c}_1} \mathrm{S}_3$, $h_j(Y) = c_1 Y_1$, $N_j = Y_1$.

2. The second-order reaction
   $\mathrm{S}_1 + \mathrm{S}_2 \xrightarrow{\mathrm{c}_2} \mathrm{S}_4$, $h_j(Y) = c_2 Y_1 Y_2$, $N_j =$min$\{Y_1, Y_2\}$.

3. The homodimer formation ($Y_1 \geq 2$)
   $\mathrm{S}_1 + \mathrm{S}_1 \xrightarrow{\mathrm{c}_3} \mathrm{S}_5$, $h_j(Y) = \frac{1}{2} c_3 Y_1 (Y_1 - 1)$, $N_j =$min$\lceil \frac{1}{2} Y_1 \rceil$.

The above equations, c_j denotes jth reaction rate and Y_i $i = 0, \ldots, n$ represents the numbers of molecules of each sprecies S_i. Whereas in the previous method, there is a single reaction to overcome the negativity problem of the Poisson distribution, this approach handle two reactions simultaneously under the condition

$$0 \leq \frac{h_j(Y) + h_l(Y)}{N_i} \leq 1, \tag{2.7}$$

where $h_j(Y)$ and $h_l(Y)$ are the propensity functions of reaction channels R_j and R_l, respectively. That is, $h_j(Y) = N_j \frac{h_j(Y)}{N_j}$ and $h_l(Y) = N_l \frac{h_l(Y)}{N_l}$.

The sampling technique for the reaction number of channels R_j and R_k is following.

1. A sample value k_{jk} for the total reaction number of R_j and R_k is generated from binomial distribution $Bin(N_i, \frac{h_j(Y)+h_l(Y)}{N_i}\tau)$, where $N_i = min\{N_j, N_l\} \neq 0$.
2. A sample value k_j for the number of reaction R_j is generated from binomial distribution $Bin(k_{jl}, \frac{h_j(Y)}{h_j(Y)+h_l(Y)})$.
3. The reaction number of R_l is obtained from $k_l = k_{jl} - k_j$ [19],[17].

Now, the method can be defined as follows. Firstly, a step size τ choosen from the τ-selection process Equation (2.13) satisfying Equation (2.6) for each reaction channel and for a given error control parameter ϵ. Then, a sample value k_j is generated from the binomial distribution $Bin(N_j, \frac{h_j(Y)\tau}{N_j})$ for $j = 1, \ldots, r$. When the reactant species undergo two or more reaction channels, the simultaneous reaction step size Equation (2.7) is applied and sampling technique is used for these reaction channels. Finally, the system is updated by [19]

$$Y(t+\tau) = Y(t) + \sum_{j=1}^{M} v_j k_j.$$

As a result, both binomial τ-leap methods ([4] and [19] allow to solve negativity problem of the Poisson τ-leap method and to advance in the computational efficiency of the algorithm owing to the larger step sizes.

2.2.5 Modified Poisson τ-leap Method

The developed version of the Poisson τ-leap method is the modified Poisson τ-leap method which aims to avoid the problem of negative populations and to obtain more accurate results than the basic Poisson τ-leap approach. In this method, there is no acceptance of the any time step τ bringing about the negativity. Until obtaining no negative populations, smaller values of τ reduced by a factor is taken. The algorithm can be explained by the following steps.

1. Calculate all the propensity functions and their sum

$$h_0(Y) = \sum_{j=1}^{r} h_j(Y)$$

   in state Y at time t.

2. Describe the set of critical reactions R_j in each τ such that currently reaction channels R_j for which $h_j(Y) > 0$ and $L_j \leq n_c (j = 1, \ldots, r)$. Hereby, $h_j(Y)$ shows the hazard function for the j^{th} reaction and $L_j = min_{v_{ij}<0} \lceil \frac{Y_i}{|v_{ij}|} \rceil$ as emphasized in the

Binomial τ-leap method. v_{ij} indicates the number of molecules $Y_i(t)$ of all substrates at time t and n_c shows the critical value setting the performance of the method anywhere between the original Poisson τ-leap method ($n_c = 0$) and the exact SSA ($n_c = \infty$).

3. Calculate the largest time step τ' regarding as the Equation (2.13) by using only the non-critical reactions j'. If there are critical reactions, all critical reactions, then replace τ' by ∞, i.e, $\tau' = \infty$.

4. If the value of τ' selected in the step 2 is less than a small multiple, like 10, of $\frac{1}{h_0(Y)}$, then the leap is declined and the exact SSA is used to get admissable number of times, like 100, before again trying a τ-leap method. In addition to these, if τ' is larger than chosen small multiple of $\frac{1}{h_0(Y)}$, then approve it and continue to step 5. Otherwise, stop and turn back inital step.

5. Evaluate the sum of the hazards of the critical reactions $h_0^c(Y)$. Generate τ'', $\tau'' \sim$ Exp($h_0^c(Y)$), with mean $\frac{1}{h_0^c(Y)}$.

6. a. If $\tau' < \tau''$: Take $\tau = \tau'$. Set $k_j = 0$ for all the critical reactions k_j. On the other hand, generate $k_j \sim$ Poi($h_j(Y)\tau$), for the non-critical reactions R_j.

   b. If $\tau'' \leq \tau'$: Take $\tau = \tau''$. Produce j_c as a sample of the integer random variable with probabilities $\frac{h_j(Y)}{h_0^c(Y)}$, where j indicates the index of critical reactions on j. Then, set $k_{j_c} = 1$ and set $k_j = 0$ for all the other critical reactions. Generate k_j sampled from Poisson distribution with mean $h_j(Y)\tau$ for the noncritical reactions R_j.

7. Update the time t and the state Y by $t := t + \tau$ and $Y := Y + \sum_{j=1}^{r} k_j v_{ij}$, respectively.

8. Undo step 7 in case that any component of Y is negative, replace τ' by $\frac{\tau'}{2}$ and go back to step 6. Otherwise, stop and turn back the initial step.

9. Save (t, Y) if desired. Return to step 1, or else stop [14], [11],[2].

2.2.6 τ-leap Method with an Improved Leap-size Selection (The new τ-selection procedure)

Due to the previously introduced methods, Poisson τ-leap (Section 2.2.1), estimated midpoint (Section 2.2.3), the following inequality should be satisfied.

$$h_j(Y + \lambda(Y)) - h_j(Y) \leq \epsilon h_0(Y), \tag{2.8}$$

where $h_j(Y)$'s $(j = 1, \ldots, r)$ are the hazard functions of each reaction j and $h_0(Y)$ is the sum $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$. Here, $\lambda(Y)$ shows the net change of state Y and ϵ denotes the error control parameter ($0 < \epsilon < 1$). Also, the largest value of τ is found as

$$\tau = min_{j\in[1,r]} \left\{ \frac{\epsilon h_0(Y)}{|\sum_{i=1}^{n} \xi_i(Y) b_{ji}|} \right\}$$

However, the approval of this value has a condition. That is, if it is less than a few multiplier of $\frac{1}{h_0(Y)}$, then it would be better to give up leap and then to use SSA. In this approach, [14] the suggestion is that $h_0(Y)$ term on the right hand side in the Equation (2.4) is replaced by $h_j(Y)$ to have more reasonable leap condition. In this case, if $h_j(Y)$ approaches to zero, this can be a problem. A new τ-selection procedure enables to determine the largest value of τ with enlarging the accuracy and without slowing down the simulation under the leap condition. Previously, the calculation of the total change is in the following form

$$f_{jj'}(Y) = \sum_{i=1}^{n} \frac{\partial h_j(Y)}{\partial Y_i} v_{ij'}, (j, j' = 1, \ldots, r). \tag{2.9}$$

A function $f_{jj'}$ is a hazard of reaction j as a consequence of the occurence of reaction j'. Hereby, n denotes the total number of substratees and $v_{ij'}$ indicates the net change of substrate i ($i = 1, \ldots, n$) as a consequence of the j'th reaction. Also, v_j or $v_{j'}$ is a n-dimensional vector. As obtained before in Section 2.2.1, in Equation (2.3), after expansion of a first-order Taylor expansion, the following equality is found.

$$\begin{aligned} \Delta h_j(Y) &= h_j(Y + \lambda(\bar{Y}, \tau)) - h_j(Y) \\ &\approx \sum_{i=1}^{n} \tau \left(\frac{\partial}{\partial Y_i} h_j(Y) \right) \sum_{j'=1}^{r} h_{j'}(Y) v_{ij'}, \end{aligned} \tag{2.10}$$

where $\lambda(\bar{Y}, \tau)$ stands for the expected change in state at time step τ. Then, as k_j is generated from a Poisson random variable with parameter $h_j(Y)\tau$, i.e., $k_j \sim$ Poi($h_j(Y)\tau$), and $\lambda(\bar{Y}, \tau) = \tau \sum_{j=1}^{r} h_j(Y) v_j$, in which k_j is the number of occurence of the jth reaction and the following approximation can be reached.

$$\lambda(\bar{Y}, \tau) \approx \sum_{j=1}^{r} Poi(h_j(Y)\tau) V_j. \tag{2.11}$$

By inserting Equation (2.10) into Equation (2.11), interchanging the two summations and employing the approximation of $\lambda(\bar{Y}, \tau)$, $\Delta h_j(Y)$ is obtained as follows

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'}(Y) Poi(h_{j'}(Y)\tau). \tag{2.12}$$

Here, the mean and variance of $\Delta h_j(Y)$ can be written by $\sum_{j'=1}^{r} f_{jj'}(Y) \times E(Poi(h_{j'}(Y)\tau))$ and $\sum_{j'=1}^{r} f_{jj'}(Y) \times Var(Poi(h_{j'}(Y)\tau))$, in order. Using these, the following expression can be derived.

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}(Y) h_j(Y)\tau \equiv \mu_j(Y)\tau$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}^2(Y) h_j(Y)\tau \equiv \sigma_j^2(Y)\tau$$

for $j = 1, \ldots, r$.

As a result, the largest acceptable τ is chosen as

$$\tau = min_j \left\{ \frac{\epsilon h_0(Y)}{|\mu_j(Y)|}, \frac{\epsilon^2 h_0^2(Y)}{\sigma_j^2(Y)} \right\} \tag{2.13}$$

since $|\mu_j(Y)\tau| \leq \epsilon h_0(Y)$ and $\sqrt{\sigma_j(Y)\tau} \leq \epsilon h_0(Y)$ and using

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm \sqrt{Var(\Delta h_j(Y))}. \tag{2.14}$$

At the final step, similar to the other methods, update the system by replacing Y and t with $Y + \lambda(Y)$ and $t + \tau$, respectively.

Unlike the previous Poisson τ-leap method in the Equation (2.4), this new approach provides extra state to choose the value of τ. However, in both case, there is no situation such that one is more accurate than the other. When to use one of them, the computationally simplier one can be chosen [7],[17].

2.2.7 Approximate Gillespie Algorithm

Lastly, one of the alternative approaches of the poisson τ-leap method is the approximate Gillespie algorithm in the literature. Basically, the approximate Gillespie algorithm [17], which depends on the extension of the exact Gillespie method. It states that k numbers of reactions, generated from the Gamma distribution with a parameter $\sum_{j=1}^{r} h_j(Y)$, where each of them occurs in an exponential time step t, is performed rather than a single reaction at a time. Then, it can be demonstrated that $\tau \sim \Gamma(k, h_0(Y))$, where τ presents the time interval of k reactions in the total hazard, $h_0(Y)$, $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$. In this case, the system is updated by replacing t by $t := t + \tau$ and by changing the current state Y by $Y := Y + \lambda(Y)$, where the net change in the state is found via $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$. In this expression v_j is the net effect of the jth reaction by showing the jth row of the net effect matrix V as used previously. By this way, it is assumed that the essential time for every reaction correspons to that of Gillespie. Under this assumption, the total number of reactions during the interval τ is determined by controlling k in each time interval. For this purpose, it is initially described a k satisfying the leap condition in each time step. Then, the change in hazard function $\Delta h_j(Y)$ $(j = 1, ..., r)$ is approximated by the first order Taylor expansion in the time interval $[t, t + \tau]$ in a such way that the following equality can be obtained as performed in the poisson τ-leap approach.

$$\Delta h_j(Y) = h_j(Y + \lambda(\bar{Y}, \tau)) - h_j(Y) \approx \lambda(\bar{Y}, \tau) h_j(Y) = \sum_{i=1}^{n} \lambda(\bar{Y}, \tau) \frac{\partial h_{ij}(Y)}{\partial Y_i} \tag{2.15}$$

in which the expected change in the state by regarding k simultaneous reaction is computed by

$$\lambda(\bar{Y}, \tau) = Y(t + \tau) - Y(t) = \sum_{j=1}^{r} k_j \nu_j.$$

In the above expression, k_j shows the number of times of the jth reaction and ν_j is the net effect of the jth reaction by denoting the jth row of the net effect matrix V as before. Hence, by using a gamma distribution, we can show $\tau \sim \Gamma(k, h_0(Y))$ where $k = E(\tau)h_0(Y)$. In this expression, $E(\tau)$ illustrates the expected τ on average.

Then, by inserting this k into Equation (2.15), the approximation of

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} f_{jj'}(Y)\tau h_0(Y),$$

is acquired where the total change in hazard of reaction j' is described in terms of $f_{jj'}$ via

$$f_{jj'} = \sum_{i=1} \frac{\partial h_j(Y)}{\partial Y_i}\nu_{ij}$$

for the execution of the reaction j'. Finally, in order to obtain the confidence interval, the following expression is written as

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm \sqrt{Var(\Delta h_j(Y))}, \tag{2.16}$$

where

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}(Y)E(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}(Y) \tag{2.17}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}^2(Y)Var(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}^2(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}^2(Y). \tag{2.18}$$

By substituting Equation (2.17) and (2.18) into the required leap condition Equation (2.2), the below expression can be found

$$|k| \sum_{j'=1}^{r} f_{jj'}(Y) \leq \epsilon h_0(Y)$$

and

$$\sqrt{\big(k \sum_{j'=1}^{r} f_{jj'}^2(Y)\big)} \leq \epsilon h_0(Y).$$

Accordingly, the optimal k is computed from

$$k = \min_j \left\lceil \frac{\epsilon h_0(Y)}{\left|\sum_{j'=1}^{r} f_{jj'}(Y)\right|}, \frac{\epsilon^2 h_0^2(Y)}{\left|\sum_{j'=1}^{r} f_{jj'}^2(Y)\right|} \right\rceil. \tag{2.19}$$

Indeed, mostly, inserting the distributions feature of k and τ into the leap condition and finding a conservative confidence interval has been derived for the poisson distribution too [2].

But in these studies, the confidence intervals are constructed one-sided and by taking a fixed significance level α which roughly sets the tabulated value to 1. Furthermore, they generate large intervals which decreases the accuracy of the approximate algorithms. Hereby, the following part introduces this study with confidence intervals so that we can produce accurate results regarding previous studies. The presentation of k via a confidence interval has not been yet to the best of our knowledge. By means of this extension, we consider to add a significance level to our analysis. Furthermore, in order to improve the accuracy of k, we present to apply sufficient statistic in the derivation of k. Specifically, we use the maximum likelihood estimator approach in the derivation of parameter. Then, we apply other alternative methods in the inference of k still within the extension of k via interval via the confidence interval. In addition to poisson distribution, we also use gamma distribution under the mentioned estimators. Therefore, we find an alternative confidence interval which is different from the intervals in the literature for the parameter k with a high accuracy.

CHAPTER 3

EXTENSION OF LEAP CONDITION

In this thesis, firstly, we use the maximum likelihood estimator, method of moment estimator and Bayesian approach to obtain an acceptable value of k which can give more accurate and more narrower result. This process is based on the advantage of the sufficient statistic in inference of parameters via the Rao-Blackwell theorem [1]. This theorem says that if the parameter is replaced by the sufficient statistics, the expected variance is smaller than the variance without the sufficient statistic while we can still keep the unbiased of the parameter. Moreover, we extend the leap condition by using higher order truncated Taylor expansion whose original estimators are found under the first order. Finally, we present alternative k interval that can have higher accuracy by with a controllable significance α.

3.1 Estimation of Model Parameters Using via Maximum Likelihood Estimator (MLE)

In order to get the maximum likelihood estimator of our parameter k, we can begin to describe the likelihood function for the value of k, denoted by $L(k)$, under the Poisson distribution, i.e, $\tau \sim Poi(k)$.

$$L(k) = \prod_{i=1}^{n} f(\tau; k) = \frac{e^{-nk} k^{\sum_{i=1}^{n} \tau_i}}{\prod_{i=1}^{n} \tau_i!},$$

where $f(\tau; k)$ is probability density function with random sample τ_i's and unknown parameter k from Poisson distribution. The natural logarithm of $L(k)$ is given by

$$ln(L(k)) = -nk + \sum_{i=1}^{n} \tau_i ln(k) - ln(\prod_{i=1}^{n} \tau_i!).$$

To find the maximum, we solve the following equation.

$$\frac{d}{dk} ln(L(k)) = -n + \sum_{i=1}^{n} \frac{\tau_i}{n} = 0$$

From here, it is reached that $\hat{k} = \frac{\sum_{i=1}^{n} \tau_i}{n} = \frac{\tau}{n}$, where the notation of hat, $\hat{}$, denotes that it is obtained from MLE.

Moreover, under the Gamma distribution with parameters k and $h_0(Y)$, that is, $\tau \sim \Gamma(k, h_0(Y))$, the likelihood function can be written by

$$L(k, h_0(Y)) = \frac{\left(\prod_{i=1}^{n} \tau_i\right)^{h_0(Y)-1} e^{-\sum_{i=1}^{n} \frac{\tau_i}{n}}}{k^{nh_0(Y)}\left(\Gamma(h_0(Y))^n\right)},$$

where the probability density function $f(\tau, k, h_0(Y))$ is shown as

$$f(\tau, k, h_0(Y)) = \frac{\tau^{h_0(Y)-1} e^{-\tau/k}}{k^{h_0(Y)\Gamma(h_0(Y))}}.$$

Then, the log-likelihood funtion $ln(L(k, h_0(Y)))$ is presented via

$$ln(L(k, h_0(Y))) = -nh_0(Y)ln(k) - nln(\Gamma(h_0(Y))) + (h_0(Y) - 1)ln(\prod_{i=1}^{n} \tau_i) - \sum_{i=1}^{n} \frac{\tau_i}{n}.$$

Later, the partial derivatives of $ln(L(k, h_0(Y)))$ are found as

$$\frac{\partial ln(L(k, h_0(Y)))}{\partial k} = -\frac{nh_0(Y)}{k} + \sum_{i=1}^{n} \frac{\tau_i}{k^2} \tag{3.1}$$

After solving Equations (3.1), $\hat{k}$ is attained as $\hat{k} = \frac{\sum_{i=1}^{n} \tau_i}{nh_0(Y)} = \frac{\tau}{nh_0(Y)}$.

3.1.1 Confidence Intervals with Poisson Distribution

As mentioned previously, MLE method can be used to infer the model parameters of approximate SSA. These parameters are k and τ, defined in terms of k. Theoretically, it is known that the MLE approach will always give sufficient estimate if it exists. Hence, while $\tau \sim Poi(k)$, the MLE of k is found as $\frac{\tau}{n}$. Then, by inserting it into $\Delta h_j(Y)$, the following expression is obtained

$$\Delta h_j(Y) = \sum_{j=1} f_{jj'}(Y)\frac{\tau}{n}.$$

With the value of $z = 1$, similar to Equation (2.17), the equality

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm \sqrt{Var(\Delta h_j(Y))}$$

is attained.

Since $\tau \sim Poi(k)$, the mean of the value τ is $E(\tau) = k$ and the variance of the value τ is found as $Var(\tau) = k$. Thus, the following equalities for the approximate values of $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are represented, respectively.

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}(Y)E(\frac{\tau}{n}) = \sum_{j'=1}^{r} f_{jj'}(Y)\frac{k}{n} = k\sum_{j'=1}^{r} \frac{f_{jj'}(Y)}{n}. \tag{3.2}$$

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}^2(Y)Var(\frac{\tau}{n}) = \sum_{j'=1}^{r} \frac{f_{jj'}^2(Y)}{n^2}k = k\sum_{j'=1}^{r} \frac{f_{jj'}^2(Y)}{n^2}. \tag{3.3}$$

After substituting Equation (3.2) and (3.3) into the required leap condition such that $|k\sum_{j'=1}^{r}\frac{f_{jj'}(Y)}{n}| \leq \epsilon h_0(Y)$ and $\sqrt{k\sum_{j'=1}^{r}\frac{f_{jj'}^2(Y)}{n^2}} \leq \epsilon h_0(Y)$, a suitable k can be derived as

$$k = \min_j \left\lceil \frac{\epsilon h_0(Y) n}{|\sum_{j'=1}^{r} f_{jj'}(Y)|}, \frac{\epsilon^2 h_0^2(Y) n^2}{|\sum_{j'=1}^{r} f_{jj'}^2(Y)|} \right\rceil \tag{3.4}$$

It can be observed that the value of $\frac{\epsilon h_0(Y) n}{\sum_{j'=1}^{r} f_{jj'}^2(Y)}$ can be smaller than the value of $\frac{\epsilon^2 h_0^2(Y) n^2}{\sum_{j'=1}^{r} f_{jj'}^2(Y)}$.

Furthermore, in order to find an admissable k, the confidence interval can be constructed for the value of k. Under large number of molecules in the collection of hazard function, it is known that the poisson distribution converges to the normal distribution as a consequence of the central limit theorem [1]. Hereby, in the derivation of the confidence interval for k, the poisson distributed random variable can be tabulated by the normal distribution. Accordingly, the critical value in the calculation of the error term can be found by the standard normal distribution table value. So, $k = \tau \pm z_{\alpha/2}\sqrt{\frac{\tau}{n}}$, while $z_{\alpha/2}$ denotes the tabulated normal value for the significance level α. Then, substituting this expression into $\Delta h_j(Y)$, the following equality is obtained as

$$\Delta h_j(Y) = \sum_{j'=1}^{r} f_{jj'}(Y)\big(\tau \pm z_{\alpha/2}\sqrt{\frac{\tau}{n}}\big).$$

For this statement, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using the assumption of $E(\sqrt{\tau}) = \sqrt{k}$ (See Appendix B.1) via

$$\begin{aligned} E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}(Y) E(\tau \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\tau}{n}}) \\ &= \sum_{j'=1}^{r} f_{jj'}(Y)\big(E(\tau) \pm \frac{z_{\frac{\alpha}{2}}}{\tau} E(\sqrt{\tau})\big) \\ &= k \pm z_{\frac{\alpha}{2}}\frac{\sqrt{k}}{\sqrt{n}} \sum_{j'=1}^{r} f_{jj'}(Y) \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}^2(Y) Var(\tau \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\tau}{n}}) \\ &= \sum_{j'=1}^{r} f_{jj'}^2(Y)\big(Var(\tau) + \frac{z_{\alpha/2}^2}{n} Var(\sqrt{\tau}) \pm \frac{2z_{\alpha/2}}{\sqrt{n}} Cov(\sqrt{\tau}, \tau)\big) \\ &= \sum_{j'=1}^{r} f_{jj'}^2(Y) k, \end{aligned} \tag{3.6}$$

as $Var(\sqrt{\tau}) \approx 0$ and $Cov(\sqrt{\tau}, \tau) \approx 0$.

Then, by inserting Equation (3.5) into the required leap condition, the following inequality is obtained.

$$\left|\left((\sqrt{k} \pm \frac{z_{\alpha/2}}{2\sqrt{n}})^2 + \frac{z_{\alpha/2}}{2\sqrt{n}}\right) \sum_{j=1}^{r} f_{jj'}\right| \leq \epsilon h_0(Y).$$

Then, it can be found that this inequality for the value of k is

$$k \leq \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}/Y} \pm \frac{z_{\alpha/2}}{2\sqrt{n}})^2} \mp \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2.$$

Applying the same process for Equation (3.6), a suitable k can be found by [5]

$$\begin{aligned} k = \min_j \Bigg\lceil & \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)} + (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2, \\ & \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)} - (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2, \\ & \frac{\epsilon^2 h_0^2(Y)}{\sum_{j'=1}^{r} f_{jj'}^2(Y)} \Bigg\rceil \end{aligned} \tag{3.7}$$

Hereby, it can be hard to compare between previous studies. However, we can say that the MLE guarantess the condition of the Cramer Rao lower bound, when an efficient estimator for k is found [3]. In other words, it is stable with less variance [1]. As a numerically, the comparison between previous result will be discussed in Section 3.6.

3.1.2 Confidence Intervals with Gamma Distribution

Under the assumption that $\tau \sim \Gamma(k, h_0(Y))$ from the distributional assumption of the approximate Gillespie method, the MLE of k is found as $\frac{\tau}{nh_0(Y)}$ where $\tau = \sum_i^n \tau_i$. Then, with the knowledge of the value of $z = 1$, similar to Equation (2.17),

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm \sqrt{Var(\Delta h_j(Y))}.$$

As $\tau \sim \Gamma(k, h_0(Y))$, the mean of the value τ is equaled to $E(\tau) = \frac{k}{h_0(Y)}$ and the variance of the value τ is computed as $Var(\tau) = \frac{k}{h_0^2(Y)}$. Thus, the following equalities for the approximate values of $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ can be obtained, in order.

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \frac{f_{jj'}(Y)}{nh_0(Y)} E(\tau) = \frac{k}{h_0^2(Y)} \sum_{j'=1}^{r} f_{jj'}(Y) \tag{3.8}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \frac{f_{jj'}^2(Y)}{n^2 h_0^2(Y)} Var(\tau) = \frac{k}{n^2 h_0^4(Y)} \sum_{j'=1}^{r} f_{jj'}^2(Y). \tag{3.9}$$

After plugging Equation (3.8) and (3.9) into the leap condition, a suitable k can be gained as

$$k = \min_j \left\lceil \left| \frac{\epsilon h_0^3(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)} \right|, \left| \frac{\epsilon^2 h_0^6(Y) n^2}{\sum_{j'=1}^{r} f_{jj'}^2(Y)} \right| \right\rceil. \tag{3.10}$$

It can be seen that the value of $\frac{\epsilon h_0^3(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)}$ is smaller than the value of $\frac{\epsilon^2 h_0^6(Y) n^2}{\sum_{j'=1}^{r} f_{jj'}^2(Y)}$ as shown in Equation (2.19) for $0 < nh_0(Y) < 1$. This implies that k found via Equation (3.10) can be generated with more accurate result.

In addition to these, to find k, the confidence interval can be constructed. Therefore, the formula $E(k) \pm \Gamma_{\alpha/2}\sqrt{\frac{Var(k)}{n}}$ as the well-known representation of the confidence interval, gives that $k \approx \frac{k}{nh_0^2(Y)} \pm \frac{\Gamma_{\alpha/2}}{nh_0(Y)}\sqrt{\frac{k}{h_0(Y)}}$. In this inequality, $\Gamma_{\alpha/2}$ denotes the tabulated normal value for the significance level α. Whereas the tabulated form of gamma distribution is not available, we prefer to convert the gamma as a chi-square distribution such a way that the chi-square, with ν degrees of freedom, i.e, χ_ν^2 is equaivalent to Gamma(γ,β) where $\gamma = \frac{\nu}{2}$ and $\beta = 2$. Moreover, it is also known that $\frac{2X}{\beta} \sim \chi_{2\gamma}^2$ if $X \sim Gamma(\gamma, \beta)$ based on the theorem in Bain and Engelhardt [1]. Therefore while $\tau \sim \Gamma(k, h_0(Y))$, $\frac{2\tau}{h_0(Y)} \sim \chi_{(2k)}^2$, that is chi-square distribution with $2k$ degrees of freedom and $\chi_{tab,(2k),\frac{\alpha}{2}}^2$ is a tabulated value for chi-square distribution with $2k$ degrees of freedom with significance level α . If we insert this expression into the confidence interval for k and thereby, k, can be defined as below:

$$E(k) - \chi_{tab,(2k),\frac{\alpha}{2}}^2 \sqrt{\frac{Var(k)}{n}}$$

and

$$E(k) + \chi_{tab,(2k),(1-\frac{\alpha}{2})}^2 \sqrt{\frac{Var(k)}{n}}$$

since χ^2 is skewed distribution and the critical vaues of both left and right side of the confidence interval have distinct cut off point, i.e, not symmetric, via $\chi_{tab,(2k),\frac{\alpha}{2}}^2 = \chi_{\frac{\alpha}{2}}^2$ and $\chi_{tab,(2k),(1-\frac{\alpha}{2})}^2 = \chi_{1-\frac{\alpha}{2}}^2$, respectively. Then, inserting this expression into $\Delta h_j(Y)$, it continues as

$$\Delta h_j(Y) = \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{k}{nh_0^2(Y)} - \frac{\chi_{\alpha/2}^2}{nh_0^2(Y)}\sqrt{\frac{k}{n}}\right) \tag{3.11}$$

and

$$\Delta h_j(Y) = \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{k}{nh_0^2(Y)} + \frac{\chi_{1-\alpha/2}^2}{nh_0^2(Y)}\sqrt{\frac{k}{n}}\right). \tag{3.12}$$

Again substituting $k = \frac{\tau}{nh_0(Y)}$ into Equation (3.11) and (3.12), the following statements are reached.

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{\tau}{n^2h_0^3(Y)} - \frac{\chi_{\alpha/2}^2}{nh_0^2(Y)}\sqrt{\frac{\tau}{n^2h_0(Y)}}\right).$$

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{\tau}{n^2h_0^3(Y)} - \frac{\chi_{\alpha/2}^2}{n^2h_0^2(Y)}\sqrt{\frac{\tau}{h_0(Y)}}\right)$$

and

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{\tau}{n^2 h_0^3(Y)} + \frac{\chi^2_{1-\alpha/2}}{n h_0^2(Y)}\sqrt{\frac{\tau}{n^2 h_0(Y)}}\right).$$
$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{\tau}{n^2 h_0^3(Y)} + \frac{\chi^2_{1-\alpha/2}}{n^2 h_0^2(Y)}\sqrt{\frac{\tau}{h_0(Y)}}\right).$$

For this statement, the derivations of the values of $E(\sqrt{\tau})$ and $Var(\sqrt{\tau})$ are evaluated by using the assumption of $E(\sqrt{\tau}) = \sqrt{\frac{k}{h_0(Y)}}$ and $Var(\sqrt{\tau}) = \sqrt{\frac{k}{h_0^2(Y)}}$ (See Appendix B.2) by

$$\begin{aligned} E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{E(\tau)}{n^2 h_0^3(Y)} - \chi^2_{\alpha/2}\frac{E(\sqrt{\tau})}{n^2 h_0^2(Y)\sqrt{h_0(Y)}}\right) \\ &= \sum_{j'=1}^{r} f_{jj'}(Y)\left(\frac{k}{n^2 h_0^4(Y)} - \chi^2_{\alpha/2}\frac{\sqrt{k}}{n^2 h_0^2(Y)\sqrt{h_0^2(Y)}}\right) \\ &= \left(\frac{k}{n^2 h_0^4(Y)} - \chi^2_{\alpha/2}\frac{\sqrt{k}}{n^2 h_0^3(Y)}\right)\sum_{j'=1}^{r} f_{jj'}(Y) \end{aligned} \tag{3.13}$$

and

$$E(\Delta h_j(Y)) \approx \left(\frac{k}{n^2 h_0^4(Y)} + \chi^2_{1-\alpha/2}\frac{\sqrt{k}}{n^2 h_0^3(Y)}\right)\sum_{j'=1}^{r} f_{jj'}(Y), \tag{3.14}$$

$$\begin{aligned} Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f^2_{jj'}(Y)\left(\frac{Var(\tau)}{n^4 h_0^6(Y)} - (\chi^2_{\alpha/2})^2\frac{Var(\sqrt{\tau})}{n^4 h_0^5(Y)}\right) \\ &= \sum_{j'=1}^{r} f^2_{jj'}(Y)\left(\frac{k}{n^4 h_0^8(Y)} - (\chi^2_{\alpha/2})^2\frac{\sqrt{k}}{n^4 h_0^6(Y)}\right) \\ &= \left(\frac{k}{n^4 h_0^8(Y)} - (\chi^2_{\alpha/2})^2\frac{\sqrt{k}}{n^4 h_0^6(Y)}\right)\sum_{j'=1}^{r} f^2_{jj'}(Y) \end{aligned} \tag{3.15}$$

and

$$Var(\Delta h_j(Y)) \approx \left(\frac{k}{n^4 h_0^8(Y)} + (\chi^2_{1-\alpha/2})^2\frac{\sqrt{k}}{n^4 h_0^6(Y)}\right)\sum_{j'=1}^{r} f^2_{jj'}(Y). \tag{3.16}$$

Then, by embedding Equation (3.13) into the required leap condition, the subsequent inequal-

ity is derived.

$$\left|\left(\frac{k}{n^2h_0^4(Y)}-\chi^2_{\alpha/2}\frac{\sqrt{k}}{n^2h_0^3(Y)}\right)\sum_{j'=1}^{r}f_{jj'}(Y)\right|\leq\epsilon h_0(Y)$$

$$\left|\left(\frac{k}{n^2h_0^4(Y)}-\chi^2_{\alpha/2}\frac{\sqrt{k}}{n^2h_0^3(Y)}+\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}-\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}\right)\sum_{j'=1}^{r}f_{jj'}(Y)\right|\leq\epsilon h_0(Y)$$

$$\left|\left(\left(\frac{\sqrt{k}}{nh_0^2(Y)}-\frac{\chi^2_{\alpha/2}}{2nh_0(Y)}\right)^2-\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}\right)\sum_{j'=1}^{r}f_{jj'}(Y)\right|\leq\epsilon h_0(Y)$$

$$\left|\left(\left(\frac{\sqrt{k}}{nh_0^2(Y)}-\frac{\chi_{\alpha/2}}{2nh_0(Y)}\right)^2-\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}\right)\right|\leq\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f_{jj'}(Y)|}.$$

Then , this inequality for the value of k can be found as

$$k\leq\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f_{jj'}(Y)|}+\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}}+\frac{\chi^2_{\alpha/2}}{2nh_0(Y)}\right)^2n^2h_0^4(Y).$$

Applying the same process for Equation (3.14), Equation (3.15) and Equation (3.16), a suitable k can be computed by

$$\begin{aligned}k\leq\min_j\Bigg\lfloor&\left|\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f_{jj'}(Y)|}+\frac{(\chi^2_{\alpha/2})^2}{4n^2h_0^2(Y)}}+\frac{\chi^2_{\alpha/2}}{2nh_0(Y)}\right)^2n^2h_0^4(Y)\right|,\\&\left|\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f_{jj'}(Y)|}+\frac{(\chi^2_{1-\alpha/2})^2}{4n^2h_0^2(Y)}}-\frac{\chi^2_{1-\alpha/2}}{2nh_0(Y)}\right)^2n^2h_0^4(Y)\right|,\\&\left|\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f^2_{jj'}(Y)|}+\frac{(\chi^2_{\alpha/2})^4}{4n^4h_0^4(Y)}}+\frac{(\chi^2_{\alpha/2})^2}{2n^2h_0^2(Y)}\right)^2n^4h_0^8(Y)\right|,\\&\left|\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r}f^2_{jj'}(Y)|}+\frac{(\chi^2_{1-\alpha/2})^4}{4n^4h_0^4(Y)}}-\frac{(\chi^2_{1-\alpha/2})^2}{2n^2h_0^2(Y)}\right)^2n^4h_0^8(Y)\right|\Bigg\rfloor.\end{aligned}\tag{3.17}$$

3.2 Estimation of Model Parameters by Using MME

Under the Poisson distribution with parameter k, $\tau\sim Poi(k)$, with probability density function $f(\tau;k)=\frac{e^{-k}k^{\tau}}{\tau!}$, $E(\tau)=k$ and $Var(\tau)=k$. Followingly, the method of moment estimation (MME) can be denoted as

$$\mu_{Poi}=\mu=k=\frac{\sum_{i=1}^{n}\tau_i}{n}=\frac{\tau}{n},$$

where μ_{Poi} denotes for the moment for poisson distribution. Here, as $E(\tau)$ and $Var(\tau)$ are the same, it is chosen just one μ, μ^1. Thus, the value of k from by using the method of moment estimator, $\tilde{k}$, under the Poisson distribution equals to $\tilde{k} = \frac{\tau}{n}$. Actually, it is the same as the value of k from MLE under poisson distribution.

Secondly, under the Gamma distribution with parameters k and $h_0(Y)$, $\tau \sim \Gamma(k, h_0(Y))$, to find the method of moment estimation, there can be written that

$$\mu^1_{Gamma} = \mu = h_0(Y)k \tag{3.18}$$

and

$$\begin{aligned} \mu^2_{Gamma} &= \sigma^2 + \mu^2 \\ &= h_0(Y)k^2 + h_0^2(Y)k^2 \\ &= k^2 h_0(Y)(1 + h_0(Y)), \end{aligned} \tag{3.19}$$

where μ^i_{Gamma} represents the moments for gamma distribution. Then, using Equation (3.18) and Equation (3.19), MMEs are found as $\tilde{k} = \sum_{i=1}^{n} \left(\frac{\tau_i - \bar{\tau}}{n\tau} \right)^2 = \frac{(n-1)n}{\bar{\tau}} S^2$, where S is a sample variance and $\tilde{h}_0(Y) = \frac{\bar{\tau}}{\tilde{k}}$. Then also, value of k can be used as $\tilde{k} = \frac{\bar{\tau}}{h_0(Y)}$ since value of $h_0(Y)$ is known. Indeed, it is known that the moment estimator cannot give guarantee for obtaining sufficient estimator as the MLE approach. But we consider that the final expression of k can be compared with other k's found by distinct estimation methods and the best k expression can be detected by simulation studies. These studies are our future works as refered in conclusion part.

Accordingly, since the value of k from MME under poisson distribution is the same as the value of k from MLE under poisson distribution, a suitable k value based on MME can be obtained from

$$k = \min_j \left\lceil \frac{\epsilon h_0(Y) n}{|\sum_{j'=1}^{r} f_{jj'}(Y)|}, \frac{\epsilon^2 h_0^2(Y) n^2}{|\sum_{j'=1}^{r} f^2_{jj'}(Y)|} \right\rceil. \tag{3.20}$$

Clearly, MLE and MME under the poisson distribution give the same result.

Furthermore, based on the gamma distribution, the moment estimation of the value k is acquired as $k = \sum_i^n (\frac{\tau_i - \tau}{n\tau})^2 = \frac{[(n-1)n]}{\tau} S^2$, where is S is a sample variance. Thereby, the mean of $\Delta h_j(Y)$, $E(\Delta h_j(Y))$ and the variance of $\Delta h_j(Y)$, $Var(\Delta h_j(Y))$ are computed approximately as the following way.

$$\begin{aligned} E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}(Y) E(\frac{(n-1)n}{\tau} S^2) \\ &= \sum_{j'=1}^{r} f_{jj'}(Y) \frac{(n-1)n}{E(\tau)} S^2 \\ &= \frac{h_0(Y) n (n-1) S^2}{k} \sum_{j=1}^{r} f_{jj'}(Y). \end{aligned} \tag{3.21}$$

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}(Y) Var(\frac{(n-1)n}{\tau} S^2) \tag{3.22}$$
$$= \frac{h_0^2(Y) n(n-1) S^2}{k} \sum_{j'=1}^{r} f_{jj'}^2(Y).$$

Then, similar to application of MLE, after inserting Equations (3.21) and (3.22) into the required leap condition, a favorable k value can be attained from

$$k \leq \min_j \left\lceil \left| \frac{n(n-1)S^2 \sum_{j=1}^{r} f_{jj'}(Y)}{\epsilon} \right|, \left| \frac{h_0(Y) n(n-1) S^2 \sum_{j=1}^{r} f_{jj'}^2(Y)}{\epsilon} \right| \right\rceil. \tag{3.23}$$

Indeed here, although Equation (3.23) can be simulated to determine which k value gives more accurate result, we cannot directly say that the expression in Equation (3.23) is narrower than the expression of k in Equation (2.19). Because, theoretically it is not guarantee that all produces estimators are sufficient statistics as the maximum likelihood estimators. Hence, this final expression is one of the future works by evaluating its value in simulation studies. Theoretically, it is not straightforward to do this.

3.3 Confidence Intervals without MLE and MME for Poisson Distribution

Alongside with these confidence interval for the value of k, there are various approaches for the confidence intervals for the mean of the Poisson distribution. The first approach is based on the normal approximation are suggested in the study of Sahai and Khurshid (1993) [18].

Thus, according to their work, the lower and the upper bound of the value of k, denoted by k_l and k_u, can be written by the following expression.

$$\begin{aligned} k_l &= k - \frac{1}{2} + \frac{1}{2} z_{1-\alpha/2}^2 + z_{1-\alpha/2} \sqrt{k - \frac{1}{2} + \frac{1}{4} z_{1-\alpha/2}^2}, \\ k_u &= k + \frac{1}{2} + \frac{1}{2} z_{1-\alpha/2}^2 + z_{1-\alpha/2} \sqrt{k + \frac{1}{2} + \frac{1}{4} z_{1-\alpha/2}^2}. \end{aligned} \tag{3.24}$$

Hereby, similar to previous approaches, by inserting Equation (3.24) into the necessary places, $\Delta h_j(Y)$ is found by

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} f_{jj'} \left(\tau \pm \frac{1}{2} + \frac{1}{2} z_{1-\alpha/2}^2 + z_{1-\alpha/2} \sqrt{\tau \pm \frac{1}{2} + \frac{1}{4} z_{1-\alpha/2}^2} \right).$$

Then, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by the properties of $E(\tau) = k$ and

$Var(\tau) = k$. Thus,

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'} E\left(\tau \pm \frac{1}{2} + \frac{1}{2}z_{1-\alpha/2}^2 + z_{1-\alpha/2}\sqrt{\tau \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2}\right) \tag{3.25}$$

$$= \sum_{j'=1}^{r} f_{jj'}\left(E(\tau) \pm \frac{1}{2} + \frac{1}{2}z_{1-\alpha/2}^2 + z_{1-\alpha/2}E\left(\sqrt{\tau \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2}\right)\right)$$

$$= \sum_{j'=1}^{r} f_{jj'}\big(k \pm \frac{1}{2} + \frac{1}{2}z_{1-\alpha/2}^2 + z_{1-\alpha/2}E(A(\tau))\big) \tag{3.26}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} f_{jj'}^2 Var\left(\tau \pm \frac{1}{2} + \frac{1}{2}z_{1-\alpha/2}^2 + z_{1-\alpha/2}\sqrt{\tau \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2}\right)$$

$$= \sum_{j'=1}^{r} f_{jj'}^2\big(Var(\tau) + z_{1-\alpha/2}^2 Var(A(\tau)) + 2z_{1-\alpha/2}Cov(\tau, A(\tau))\big)$$

$$= \sum_{j'=1}^{r} f_{jj'}^2\big(k + z_{1-\alpha/2}^2 Var(A(\tau)) + 2z_{1-\alpha/2}Cov(\tau, A(\tau))\big). \tag{3.27}$$

In order to simplify the above expression, $A(\tau)$ is defined as $A(\tau) = \sqrt{\tau \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2}$.

Then, applying the similar process for calculating $E(\sqrt{\tau})$ (See AppendixB.2) to find $E(A(\tau))$, it can be approximated by $E(A(\tau)) \approx \sqrt{k \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2}$. Also, under the same approximations, it can be reached that $Var(A(\tau)) \approx 0$ and $Cov(\tau, A(\tau)) \approx 0$. By substituting Equation (3.25) and (3.27) into the required leap condition, the following inequalities have

$$k + z_{1-\alpha/2}^2\sqrt{k \pm \frac{1}{2} + \frac{1}{4}z_{1-\alpha/2}^2} \leq \frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} \mp \frac{1}{2} - \frac{1}{2}z_{1-\alpha/2}^2$$

$$k \leq \left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{z_{1-\alpha/2}^2}{4}} - \frac{z_{1-\alpha/2}}{2}\right)^2$$

$$\mp \frac{1}{2} + \frac{z_{1-\alpha/2}^2}{2} \tag{3.28}$$

and

$$k \leq \frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}^2(Y)|}. \tag{3.29}$$

Hence, the suitable value of k can be chosen as

$$k \leq \min_{j\in}\Bigg\lceil\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{z_{1-\alpha/2}^2}{4}} - \frac{z_{1-\alpha/2}}{2}\right)^2 - \frac{1}{2} + \frac{z_{1-\alpha/2}^2}{2},$$

$$\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{z_{1-\alpha/2}^2}{4}} - \frac{z_{1-\alpha/2}}{2}\right)^2 + \frac{1}{2} + \frac{z_{1-\alpha/2}^2}{2}, \frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}^2(Y)|}\Bigg\rceil.$$

Equation (3.30) is gained under some approximations and assumptions. Considering these issues, the solution is not general and is valid under these conditions.

Moreover, the confidence interval via the lower and upper bound of k with a normal approximation can be written more simply as [18].

$$k_l = k - z_{1-\alpha/2}\sqrt{\tau}.$$
$$k_u = k + z_{1-\alpha/2}\sqrt{\tau}.$$

Repeating the same procedure for this value k's, and as a common choosing α such that $z_{\alpha/2} = 1$ to simplify, the value of k can obtained as

$$k \leq \min_j \left\lceil \left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{1}{4}} - \frac{1}{2} \right)^2, \left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{1}{4}} + \frac{1}{2} \right)^2, \frac{\epsilon^2 h_0^2(Y)}{|\sum_{j'=1}^{r} f_{jj'}^2(Y)|} \right\rceil. \tag{3.30}$$

Lastly, an improved normal approximation for the confidence interval gives us the following k with k_l and k_u based on the study of Sahai and Khurshid (1993) [18].

$$k_l = k + \frac{2z_{\alpha/2}^2 + 1}{6} - \left(\frac{1}{2} + \sqrt{2z_{\alpha/2}^2 \big(k - \frac{1}{2} + \frac{2z_{\alpha/2}^2 + 1}{18}\big)} \right).$$

$$k_u = k + \frac{2z_{\alpha/2}^2 + 1}{6} + \left(\frac{1}{2} + \sqrt{2z_{\alpha/2}^2 \big(k + \frac{1}{2} + \frac{2z_{\alpha/2}^2 + 1}{18}\big)} \right).$$

Finally, by implemeting a similar process as before and using $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$, the subsequents can be found that

$$k \mp E(B(\tau)) \leq \frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} - z_{\alpha/2}^2 \frac{1}{3} - \frac{1}{6} \pm \frac{1}{2} \tag{3.31}$$

and

$$k + z_{\alpha/2}^2 Var(B(\tau)) + 2z_{\alpha/2} Cov(\tau, B(\tau)) \leq \frac{\epsilon^2 h_0^2(Y)}{\sum_{j=1}^{r} f_{jj'}^2(Y)}, \tag{3.32}$$

where $B(\tau) = \sqrt{\tau \pm \frac{1}{2} + \frac{z_{\alpha/2}^2 + 2}{18}}$.

Then, like the previous studies, under similar assumption and approximations, choosing α such that $z_{\alpha/2} = 1$, inequalities (3.31) and (3.32) can imply

$$k + \sqrt{k} \leq \frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)} - \frac{1}{3} - \frac{1}{6} \pm \frac{1}{2},$$
$$k \leq \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)} - \frac{1}{3} + \frac{1}{12} \pm \frac{1}{2}} - \frac{1}{2} \right) \tag{3.33}$$

and

$$k \leq \frac{\epsilon^2 h_0^2(Y)}{\sum_{j=1}^{r} f_{jj'}^2(Y)}. \tag{3.34}$$

Therefore, an appropriate value of k can be presented as

$$k \leq \min_j \left\lceil \left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} - \frac{9}{12}} - \frac{1}{2} \right), \left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} + \frac{3}{12}} - \frac{1}{2} \right), \right.$$
$$\left. \frac{\epsilon^2 h_0^2(Y)}{|\sum_{j=1}^{r} f_{jj'}^2(Y)|} \right\rceil. \tag{3.35}$$

Hereby, it can be easily inferred that the value of $\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|} - \frac{9}{12}} - \frac{1}{2} \right)$ is smaller than the value of $\frac{\epsilon h_0(Y)}{\left|\sum_{j'=1}^{r} f_{jj'}(Y)\right|}$. Moreover, the other outputs can be compared directly by simulation. Actually, it is not theoretically straightforward because of square root.

3.4 Estimation of Model Parameters via Bayesian Approach

In addition to these confidence intervals obtained by MLE and MME, it can be possible to obtain appropriate value of k by using the bayesian estimator. For this purpose, we take the prior distribution of our parameters as $\tau \sim Poi(k)$ and $k \sim \Gamma(\alpha, \beta)$, where α and β are the given parameters [1]. Then, the conditional posterior of $\tau|k$ is derived as the gamma distribution due to the conjugate relation between the poisson and the gamma. Hence, the expectation of the conditional posterior is found as

$$E(\tau|k) = \frac{\sum \tau_i + \beta}{n + \frac{1}{\alpha}}. \tag{3.36}$$

By inserting Equation (3.36) into the function $\Delta h_j(Y)$, the following expression is obtained.

$$\Delta h_j(Y) = \sum_{j'=1} f_{jj'}(Y) \frac{\sum \tau_i + \beta}{n + \frac{1}{\alpha}}.$$

Then, it can be seen the following equality by taking $E(\tau) = k$ and $Var(\tau) = k$ as the property of the Poisson distribution $\tau \sim Poi(k)$.

$$E(\Delta h_j(Y)) \approx \sum f_{jj'}(Y) \frac{E(\tau) + \beta}{n + \frac{1}{\alpha}} \tag{3.37}$$
$$= \frac{k + \beta}{n + \frac{1}{\alpha}} \sum f_{jj'}(Y)$$

and

$$Var(\Delta h_j(Y)) \approx \sum \frac{f_{jj'}^2(Y)}{\left(n + \frac{1}{\alpha}\right)^2} Var(\tau) \tag{3.38}$$
$$= \frac{k}{\left(n + \frac{1}{\alpha}\right)^2} \sum f_{jj'}^2(Y).$$

Later, similar to previous derivations, by substituting them into the leap condition, the coming inequalities can be derived as

$$k \leq \frac{(n+\frac{1}{\alpha})\epsilon h_0(Y)}{|\sum f_{jj'}(Y)|} - \beta,$$
$$k \leq \frac{(n+\frac{1}{\alpha})^2\epsilon^2 h_0^2(Y)}{|\sum f_{jj'}^2(Y)|}.$$

As a result, an appropriate value of k can be reached by the inequality below.

$$k \leq \min_j \left\lceil \frac{(n+\frac{1}{\alpha})\epsilon h_0(Y)}{|\sum f_{jj'}(Y)|} - \beta, \frac{(n+\frac{1}{\alpha})^2\epsilon^2 h_0^2(Y)}{|\sum f_{jj'}^2(Y)|} \right\rceil. \tag{3.39}$$

By this way, it can generate more flexible and narrower confidence intervals for k due to the controlable significance level α and appropriate estimation techniques, respectively.

3.5 Extension of Leap Condition via 2^{nd} and 3^{rd} Order Truncated Taylor Expansion

Previously, using the first order Taylor expansion of hazard function $h_0(Y)$, the novel derivations that we introduce have been presented. Now, in this part, $h_0(Y)$ is expanded by using 2^{nd} and 3^{rd} order truncated Taylor formula. This expansion provides more information about the distribution. In other words, 2^{nd} and 3^{rd} order expansion can supply the knowledge of the variance and covariance, respectively. Actually, these are also known as 2^{nd} and 3^{rd} moment. Moreover, the obtained confidence intervals are narrower than the results from the previous studies. So that we can obtain a more accurate k by forming confidence interval with parameters of estimation deriving MLE and MME under Poisson and Gamma distribution due to Rao-Blackwell theorem [1], [3], seperately.

3.5.1 Extension of Leap Condition by 2^{nd} Order Taylor Expansion

We consider to apply 2^{nd} order truncated Taylor expansion, rather than using 1^{st} order as the following way.

$$h_j(Y+\bar{\lambda}(Y,\tau)) \approx h_j(Y) + \bar{\lambda}(Y,\tau)\nabla h_j(Y) + \frac{1}{2}\bar{\lambda}(Y,\tau)H(Y)\bar{\lambda}^T(Y,\tau).$$
$$h_j(Y+\bar{\lambda}(Y,\tau)) - h_j(Y) \approx \bar{\lambda}(Y,\tau)\nabla h_j(Y) + \frac{1}{2}\bar{\lambda}(Y,\tau)H(Y)\bar{\lambda}^T(Y,\tau).$$
$$\Delta h_j(Y) \approx \bar{\lambda}(Y,\tau)\nabla h_j(Y) + \frac{1}{2}\bar{\lambda}(Y,\tau)H(Y)\bar{\lambda}^T(Y,\tau), \tag{3.40}$$

where $\nabla h_j(Y)$ is a gradient of $h_j(Y)$ such that $\nabla h_j(Y) = \sum_{i=1}^n \frac{\partial h_j(Y)}{\partial Y_i}$, $H(Y)$ denotes a hessian matrix of $h_j(Y)$ such that $H(Y) = \sum_{i,l=1}^n \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l}$ and $\bar{\lambda}(Y,t) = Y(t+\tau) - Y(t) =$

$\sum_{j=1}^{r} k_j \nu_j$. Then, $\Delta h_j(Y)$ can be approximated by the following way.

$$\Delta h_j(Y) \approx \bar{\lambda}(Y,\tau)\nabla h_j(Y) + \frac{1}{2}\bar{\lambda}(Y,\tau)H(Y)\bar{\lambda}^T(Y,\tau).$$
$$\Delta h_j(Y) \approx \sum_{i=1}^{n} \bar{\lambda_i}(Y,\tau)\frac{\partial h_{ij}(Y)}{\partial Y_i} + \frac{1}{2}\sum_{i,l=1}^{n} \bar{\lambda_i}^T(Y,\tau)\frac{\partial^2 h_{ij}(Y)}{\partial Y_i \partial Y_l}\bar{\lambda_i}(Y,\tau), \tag{3.41}$$

where $\bar{\lambda_i}(Y,\tau) = \sum_{j=1}^{r} k_j v_{ji}$ and $\bar{\lambda_i}^T(Y,\tau) = \sum_{j=1}^{r} v_{ji} k_j$. Accordingly, we can obtained the subsequent statement.

$$\Delta h_j(Y) \approx k \sum_{j'=1}^{r} f_{jj'}(Y) + \frac{1}{2}k^2 \sum_{j'=1}^{r} g_{jj'}(Y), \tag{3.42}$$

where

$$f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i}$$

and

$$g_{jj'}(Y) = \sum_{l=1}^{n}\sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}.$$

Then, considering τ is generated from gamma distribution with parameters k and $h_0(Y)$, i.e., $\tau \sim \Gamma(k, h_0(Y))$, k can be written as $k = \tau.h_0(Y)$ by using the expectation. After substituting this final expression k into Equation (3.42), $\Delta h_j(Y)$ can be obtained in the following way.

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} \big(\tau h_0(Y) f_{jj'}(Y) + \frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big). \tag{3.43}$$

We have that

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm z_{\alpha/2}\sqrt{Var(\Delta h_j(Y))} \tag{3.44}$$

as seen in the previous studies in Section (2.2.7). Then, the mean of Equation (3.43), $E(\Delta h_j(Y))$, can be calculated by

$$\begin{aligned} E(\Delta h_j(Y)) &\approx E\bigg(\sum_{j'=1}^{r}\big(\tau h_0(Y) f_{jj'}(Y) + \frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big)\bigg) \\ &= \sum_{j'=1}^{r}\bigg(E(\tau) h_0(Y) f_{jj'}(Y) + E(\tau^2)\frac{h_0^2(Y) g_{jj'}(Y)}{2}\bigg). \end{aligned} \tag{3.45}$$

Since $E(\tau) = \frac{k}{h_0(Y)}$ and $E(\tau^2) = \frac{k(k+1)}{h_0^2(Y)}$,

$$\begin{aligned}
E(\Delta h_j(Y)) &\approx E\Big(\sum_{j'=1}^{r} \big(\tau h_0(Y) f_{jj'}(Y) + \frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big)\Big) \\
&= \sum_{j'=1}^{r} \left(\frac{k}{h_0(Y)} h_0(Y) f_{jj'}(Y) + \frac{k(k+1)}{h_0^2(Y)} \frac{h_0^2(Y) g_{jj'}(Y)}{2} \right) \\
&= \sum_{j'=1}^{r} \big(k f_{jj'}(Y) + \frac{k(k+1)}{2} g_{jj'}(Y)\big) \\
&= \sum_{j'=1}^{r} \left(\frac{k^2}{2} g_{jj'}(Y) + k\big(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2}\big) \right) \\
&= \sum_{j'=1}^{r} \left(\left(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 - \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 \right).
\end{aligned} \tag{3.46}$$

Then, by substituting this final expression in Equation (3.46) into the leap condition similar to the previous works, the subsequent inequality can be acquired.

$$\left| \sum_{j'=1}^{r} \left(\left(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 - \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 \right) \leq \epsilon h_0(Y) \right|. \tag{3.47}$$

After that,

$$\left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 - \epsilon h_0(Y) \leq \sum_{j'=1}^{r} \left(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 \tag{3.48}$$

and

$$\sum_{j'=1}^{r} \left(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2 \leq \epsilon h_0(Y) + \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2. \tag{3.49}$$

Using the inequality in (3.49), the following expressions are found.

$$\sqrt{\sum_{j'=1}^{r} \left(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2} \leq \sqrt{\epsilon h_0(Y) + \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2}.$$

$$\sum_{j'=1}^{r} k\sqrt{\frac{g_{jj'}(Y)}{2}} \leq \sqrt{\epsilon h_0(Y) + \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2} - \sum_{j'=1}^{r} \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}.$$

So,

$$k \leq \frac{\sqrt{\epsilon h_0(Y) + \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}} \right)^2} - \sum_{j'=1}^{r} \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}}{\sum_{j=1}^{r} \sqrt{\frac{g_{jj'}(Y)}{2}}}. \tag{3.50}$$

Moreover, by using the inequality in (3.48) and doing the same process appyling the expression in Equation (3.49), the following inequality can be reached.

$$\frac{\sqrt{\left(\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\right)^2-\epsilon h_0(Y)}-\sum_{j'=1}^{r}\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}}{\sum_{j=1}^{r}\sqrt{\frac{g_{jj'}(Y)}{2}}}\leq k. \tag{3.51}$$

In a similar way of the calculation of $E(\Delta h_j(Y))$, $Var(\Delta h_j(Y))$ is gained in the following way.

$$\begin{aligned}Var(\Delta h_j(Y)) &\approx Var\bigg(\sum_{j'=1}^{r}\big(\tau h_0(Y)f_{jj'}(Y)+\frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big)\bigg).\\ &=\sum_{j'=1}^{r}\bigg(h_0^2(Y)f_{jj'}(Y)Var(\tau)+\frac{h_0^4 g_{jj'}(Y)}{4}Var(\tau^2)\\ &+h_0^3(Y)f_{jj'}(Y)g_{jj'}(Y)Cov(\tau,\tau^2)\bigg).\end{aligned} \tag{3.52}$$

As $\tau\sim\Gamma(k,h_0(Y)$, it is known that $Var(\tau)=\frac{k}{h_0^2(Y)}$, $Var(\tau^2)=\frac{4k^3+10k^2+6k}{h_0^4(Y)}$ $=\frac{2k(k+1)(2k+3)}{h_0^4(Y)}$ and $Cov(\tau,\tau^2)=\frac{2k(k+1)}{h_0^3(Y)}$. Substituting these values into the equation of $Var(\Delta h_j(Y))$ (Equation (3.52)), the expression below is derived.

$$\begin{aligned}Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r}\bigg(h_0^2(Y)f_{jj'}(Y)\frac{k}{h_0^2(Y)}+\frac{h_0^4(Y)g_{jj'}(Y)}{4}\frac{2k(k+1)(2k+3)}{h_0^4(Y)}\\ &+h_0^3(Y)f_{jj'}(Y)g_{jj'}(Y)\frac{2k(k+1)}{h_0^3(Y)}\bigg).\\ &=\sum_{j'=1}^{r}\big(f_{jj'}^2(Y)k+\frac{g_{jj'}^2(Y)2k(k+1)(2k+3)}{4}+f_{jj'}(Y)g_{jj'}(Y)2k(k+1)\big).\end{aligned} \tag{3.53}$$

After plugging into the leap condition, it is obtained that

$$k\leq\frac{\sqrt{\epsilon^2h_0^2(Y)-A(Y)+\sum_{j'=1}^{r}B^2(Y)}-\sum_{j'=1}^{r}B(Y)}{\sum_{j=1}^{r}g_{jj'}(Y)}, \tag{3.54}$$

where $A(Y)=\sum_{j'=1}^{r}\big(2f_{jj'}(Y)g_{jj'}(Y)+f_{jj'}^2(Y)+\frac{3}{2}g_{jj'}^2(Y)\big)$ and $B(Y)=\left(\frac{\frac{5}{2}g_{jj'}(Y)+2f_{jj'}(Y)g_{jj'}(Y)}{2g_{jj'}(Y)}\right)$.

$$\begin{aligned}k=\min_{j\in[1,r]}\Bigg\lceil&\frac{\sqrt{\epsilon h_0(Y)+\left(\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\right)^2}-\sum_{j=1}^{r}\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}}{\sum_{j'=1}^{r}\sqrt{\frac{g_{jj'}(Y)}{2}}},\\ &\frac{\sqrt{\epsilon^2h_0^2(Y)-A(Y)+\sum_{j=1}^{r}B^2(Y)}-\sum_{j'=1}^{r}B(Y)}{\sum_{j'=1}^{r}g_{jj'}(Y)}\Bigg\rceil,\end{aligned} \tag{3.55}$$

where $A(Y) = \sum_{j'=1}^{r} \left(2f_{jj'}(Y)g_{jj'}(Y) + f_{jj'}^2(Y) + \frac{3}{2}g_{jj'}^2(Y)\right)$ and $B(Y) = \left(\frac{\frac{5}{2}g_{jj'}(Y)+2f_{jj'}(Y)g_{jj'}(Y)}{2g_{jj'}(Y)}\right)$. Hereby, the aim of defining $A(Y)$ and $B(Y)$ is a simplification for the writing. As defined above, the function $g_{jj'}(Y)$ comes from 2^{nd} order truncated Taylor expansion. Moreover, it can be observable that when the function $g_{jj'}(Y)$ will be close to zero, then the expression can be the same as the output from 1^{st} order Taylor expansion.

3.5.1.1 Confidence Intervals with Poisson Distribution by Using MLE

From MLE under the Poisson distribution, $\tau \sim Poi(k)$, it is known that the value of k equals to $k = \frac{\tau}{n}$ from Section 3.1.1. Then, this value is substituted into $\Delta h_j(Y)$ in Equation (3.42). Then, $\Delta h_j(Y)$ equals to

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} \left(\frac{\tau}{n} f_{jj'}(Y) + \frac{\tau^2}{2n^2} g_{jj'}(Y)\right), \tag{3.56}$$

where

$$f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i}$$

and

$$g_{jj'}(Y) = \sum_{l=1}^{n} \sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}.$$

Evaluating $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ in the following expression.

$$E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \left(E(\tau)\frac{f_{jj'}(Y)}{n} + E(\tau^2)\frac{g_{jj'}(Y)}{2n^2}\right). \tag{3.57}$$

Since $\tau \sim Poi(k)$, the equalities of $E(\tau) = k$ and $E(\tau^2) = k + k^2$ are known. Thus,

$$\begin{aligned}
E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} \left(\frac{k}{n} f_{jj'}(Y) + \frac{(k+k^2)}{2n^2} g_{jj'}(Y)\right), \\
&= \sum_{j'=1}^{r} \left(\left(\frac{f_{jj'}(Y)}{n} + \frac{g_{jj'}(Y)}{2n^2}\right)k + \frac{g_{jj'}(Y)}{2n^2}k^2\right), \\
&= \sum_{j'=1}^{r} \left(\left(\frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}}k + \frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}}\right)^2\right. \\
&\quad \left. - \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}}\right)^2\right)
\end{aligned} \tag{3.58}$$

and also, $Var(\tau) = k$, $Var(\tau^2) = k + 6k^2 + 4k^3$ and $Cov(\tau, \tau^2) = 2k^2 + k$ are considered due to Poisson distribution [1]. Then, the following equalities are obtained.

$$\begin{aligned}
Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} Var\left(\frac{\tau}{n} f_{jj'}(Y) + \frac{\tau^2}{2n^2} g_{jj'}(Y), \right) \\
&= \sum_{j'=1}^{r} \left(\frac{f_{jj'}^2(Y)}{n} Var(\tau) + \frac{g_{jj'}^2(Y)}{4n^4} Var(\tau^2) + \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^3} Cov(\tau, \tau^2)\right), \\
&= \sum_{j'=1}^{r} \left(\frac{f_{jj'}^2(Y)}{n^2} k + \frac{g_{jj'}^2(Y)}{4n^4}(k + 6k^2 + 4k^3) + \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^3}(2k^2 + k)\right), \\
&= \sum_{j'=1}^{r} \left(\frac{g_{jj'}^2(Y)}{n^4} k^3 + \left(\frac{3g_{jj'}(Y)}{2n^4} + \frac{2f_{jj'}(Y) g_{jj'}(Y)}{n^3}\right) k^2 \right. \\
&\left. + \left(\frac{f_{jj'}^2(Y)}{n^2} + \frac{g_{jj'}^2(Y)}{4n^4} + \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^3} + \right) k\right) \qquad (3.59)
\end{aligned}$$

After substituting Equation (3.58) into the leap condition, the expression becomes the following form

$$\left| \sum_{j'=1}^{r} \left(\left(\frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}} k + \frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)^2 - \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)^2 \right) \right| \leq \epsilon h_0(Y). \qquad (3.60)$$

Subsequently, the inequality below can be inferred from Equation in (3.60).

$$\left(\frac{\sqrt{\sum_{j'=1}^{r} \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)^2 - \epsilon h_0(Y)} - \sum_{j'=1}^{r} \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)}{\sum_{j'=1}^{r} \frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}}} \right) \leq k \qquad (3.61)$$

and

$$k \leq \frac{\sqrt{\epsilon h_0(Y) + \sum_{j'=1}^{r} \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)^2} - \sum_{j'=1}^{r} \left(\frac{2f_{jj'}(Y)n + g_{jj'}(Y)}{2n\sqrt{2g_{jj'}(Y)}} \right)}{\sum_{j'=1}^{r} \frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}}}. \qquad (3.62)$$

Similarly, substituting Equation (3.59) into the leap condition gives that

$$k \leq \frac{\sqrt{\epsilon^2 h_0^2(Y) - \sum_{j'=1}^{r} \left(\frac{g_{jj'}^2(Y) - g_{jj'}(Y)}{4n^4 - \frac{f_{jj'}(Y) g_{jj'}(Y) - 6f_{jj'}(Y)}{n^3}} - \frac{3f_{jj'}^2(Y)}{n^2} \right)} - \sum_{j'=1}^{r} \left(\frac{3}{2n^2} - \frac{2f_{jj'}(Y)}{n} \right)}{\sum_{j'=1}^{r} \frac{g_{jj'}(Y)}{n^2}}. \qquad (3.63)$$

Consequently, from Equation (3.62) and Equation (3.63), an admissable value of k is chosen by

$$k = \min_j \left\lceil \frac{\sqrt{\epsilon h_0(Y) + \sum_{j'=1}^{r} \left(\frac{f_{jj'}(Y) + \frac{g_{jj'}(Y)}{n^2}}{\sqrt{2g_{jj'}(Y)}} \right)^2} - \sum_{j'=1}^{r} \left(\frac{f_{jj'}(Y) + \frac{g_{jj'}(Y)}{n^2}}{\sqrt{2g_{jj'}(Y)}} \right)}{\sum_{j'=1}^{r} \frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}}}, \right.$$
$$\left. \frac{\sqrt{\epsilon^2 h_0^2(Y) + \sum_{j'=1}^{r} \left(f_{jj'}(Y) + \frac{3g_{jj'}(Y)}{4n^2} \right)^2} - \sum_{j'=1}^{r} \left(f_{jj'}(Y) + \frac{3g_{jj'}(Y)}{4n^2} \right)}{\sum_{j'=1}^{r} \frac{\sqrt{g_{jj'}(Y)}}{n}} \right\rceil. \quad (3.64)$$

It is not tractable to directly to compare Equation (3.64) to the previous values. Because there are terms with square roots and a lot of terms. Also, it is not certain the interval of these terms in Equation (3.64). For example, the functions $f_{jj'}(Y)$ and $g_{jj'}(Y)$ are consisted of hazard functions and hazards is a combination of states Y_i's and reaction rate constants c_i's. So, hazard function is prone to reduce during the reaction as the number of molecules is decrasing. Moreover, it is not clear whether the terms is bigger than 1 or not. It is just known that reaction rate constants are positive. This reason makes also difficulty in the comparison. However, it is known that MLE can give more accurate result than outputs from other estimators if it exists and the estimation 2^{nd} order truncated Taylor series expansion gives more accurate results than the estimation by the 1^{st} order Taylor expansion formula. Therefore, it can be comparible with the results in Equation (3.4), where $\Delta h_j(Y)$ is expanded by the first order Taylor formula. Accordingly, due to the underlying theoretical reason in higher order expansion, we can conclude that Equation (3.64) is narrower than Equation (3.4).

In addition to these , the confidence interval under the Poisson distribution for the value of k estimated from MLE is obtained as, previously in Section 3.1.1, $k \approx \frac{\tau}{n} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tau}{n^2}}$. Then, plugging this interval into Equation (3.42) gives the following expression.

$$\begin{aligned}
\Delta h_j(Y) &\approx \sum_{j'=1}^{r} \left(\left(\frac{\tau}{n} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tau}{n^2}} \right) f_{jj'}(Y) + \frac{1}{2} \left(\frac{\tau}{n} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tau}{n^2}} \right)^2 g_{jj'}(Y) \right), \\
&= \sum_{j'=1}^{r} \left(\frac{g_{jj'}(Y)}{2n} \tau^2 \pm z_{\alpha/2} \frac{g_{jj'}(Y)}{n^2} \tau\sqrt{\tau} + \left(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \right) \tau \right. \\
&\left. \pm z_{\alpha/2} \frac{f_{jj'}(Y)}{n} \sqrt{\tau} \right), \quad (3.65)
\end{aligned}$$

where

$$f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i}$$

and

$$g_{jj'}(Y) = \sum_{l=1}^{n} \sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}.$$

Then, the values of $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using the equalities below.

$$E(\tau\sqrt{\tau}) \approx k\sqrt{k},$$

having a similar process of obtaining $E(\sqrt{\tau})$,

$$\begin{aligned}
Var(\sqrt{\tau}) &\approx 0,\\
Var(\tau\sqrt{\tau}) &\approx k + 3k^2,\\
Cov(\tau^2, \tau\sqrt{\tau}) &\approx -k^2\sqrt{k},\\
Cov(\tau^2, \tau) &= 2k^2 + k,\\
Cov(\tau^2, \sqrt{\tau}) &= -k\sqrt{k},\\
Cov(\tau\sqrt{\tau}) &\approx 0,\\
Cov(\tau\sqrt{\tau}, \sqrt{\tau}) &\approx k,\\
Cov(\tau, \sqrt{\tau}) &\approx 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r}\Bigg(\frac{g_{jj'}(Y)}{2n}E(\tau^2) \pm z_{\alpha/2}\frac{g_{jj'}(Y)}{n^2}E(\tau\sqrt{\tau})\\
&\quad + \left(\frac{f_{jj'}(Y)}{n} + z^2_{\alpha/2}\frac{g_{jj'}(Y)}{2n^2}\right)E(\tau) \pm z_{\alpha/2}\frac{f_{jj'}(Y)}{n}E(\sqrt{\tau})\Bigg),\\
&= \sum_{j'=1}^{r}\Bigg(\frac{g_{jj'}(Y)}{2n}(k + k^2) \pm z_{\alpha/2}\frac{g_{jj'}(Y)}{n^2}(k\sqrt{k})\\
&\quad + \left(\frac{f_{jj'}(Y)}{n} + z^2_{\alpha/2}\frac{g_{jj'}(Y)}{2n^2}\right)k \pm z_{\alpha/2}\frac{f_{jj'}(Y)}{n}\sqrt{k}\Bigg) \qquad (3.66)
\end{aligned}$$

and

$$
\begin{aligned}
Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \Bigg(& \frac{g_{jj'}^2(Y)}{4n^2} Var(\tau^2) + z_{\alpha/2}^2 \frac{g_{jj'}^2(Y)}{n^4} Var(\tau\sqrt{\tau}) \\
& + \left(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \right)^2 Var(\tau) + z_{\alpha/2}^2 \frac{f_{jj'}^2(Y)}{n^2} Var(\tau) \\
& - z_{\alpha/2} \frac{g_{jj'}^2(Y)}{n^3} Cov(\tau^2, \tau\sqrt{\tau}) + \frac{g_{jj'}(Y)}{n} \big(\frac{f_{jj'}(Y)}{n} \\
& + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \big) Cov(\tau^2, \tau) \pm z_{\alpha/2} \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^2}) Cov(\tau^2, \sqrt{\tau}) \\
& + z_{\alpha/2} \frac{g_{jj'}(Y)}{n^2} \big(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \big) Cov(\tau\sqrt{\tau}, \tau) \\
& + z_{\alpha/2}^2 \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^3} Cov(\tau\sqrt{\tau}, \sqrt{\tau}) \\
& + z_{\alpha/2} \frac{f_{jj'}(Y)}{n} \big(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \big) Cov(\tau, \sqrt{\tau}) \Bigg), \\
= \sum_{j'=1}^{r} \Bigg(& \frac{g_{jj'}^2(Y)}{4n^2} (k + 6k^2 + 4k^3) + z_{\alpha/2}^2 \frac{g_{jj'}^2(Y)}{n^4} (k + 3k^2) \\
& + \left(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \right)^2 k + z_{\alpha/2}^2 \frac{f_{jj'}^2(Y)}{n^2} Var(\tau) \\
& - z_{\alpha/2} \frac{g_{jj'}^2(Y)}{n^3} k^2\sqrt{k} + \frac{g_{jj'}(Y)}{n} \big(\frac{f_{jj'}(Y)}{n} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^2} \big) (2k^2 + k) \\
& \pm z_{\alpha/2} \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^2} k\sqrt{k} + z_{\alpha/2}^2 \frac{2 f_{jj'}(Y) g_{jj'}(Y)}{n^3} k \Bigg). \qquad (3.67)
\end{aligned}
$$

From Equation (3.66) and Eqaution (3.67), by getting k alone is effortful since the equations have not integer powers. In other words, the equations have roots powers. Hence, the process of constructing confidence interval for k at the beginning and inserting Equation (3.42) is likely to be inefficient as a workload.

3.5.1.2 Confidence Intervals with Gamma Distribution by Using MLE

From MLE, it is known that the value of k equals to $k = \frac{\tau}{nh_0(Y)}$. Then, this value is inserted $\Delta h_j(Y)$ in Equation (3.42).

$$
\Delta h_j(Y) \approx \sum_{j'=1}^{r} \left(\frac{\tau}{nh_0(Y)} f_{jj'}(Y) + \frac{\tau^2}{2n^2 h_0^2(Y)} g_{jj'}(Y) \right), \qquad (3.68)
$$

where

$$
f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i}
$$

and

$$
g_{jj'}(Y) = \sum_{l=1}^{n} \sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}.
$$

Thereby,

$$
\begin{aligned}
E(\Delta h_j(Y)) &\approx E\Bigg(\sum_{j'=1}^{r}\left(\frac{\tau}{nh_0(Y)}f_{jj'}(Y)+\frac{\tau^2}{2n^2h_0^2(Y)}g_{jj'}(Y)\right)\Bigg),\\
&=\sum_{j'=1}^{r}\left(\frac{E(\tau)}{nh_0(Y)}f_{jj'}(Y)+\frac{E(\tau^2)}{2n^2h_0^2(Y)}g_{jj'}(Y)\right),\\
&=\sum_{j'=1}^{r}\left(\frac{k}{nh_0^2(Y)}f_{jj'}(Y)+\frac{k(k+1)}{2n^2h_0^4(Y)}g_{jj'}(Y)\right),\\
&=\sum_{j'=1}^{r}\left(\frac{g_{jj'}(Y)}{2n^2h_0^4(Y)}k^2+\left(\frac{f_{jj'}(Y)}{nh_0^2(Y)}+\frac{g_{jj'}(Y)}{2n^2h_0^4(Y)}\right)k\right),\\
&=\sum_{j'=1}^{r}\Bigg(\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}k+\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}\right)^2\\
&-\left(\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}\right)^2\Bigg)
\end{aligned}
\tag{3.69}
$$

and

$$
\begin{aligned}
Var(\Delta h_j(Y)) &\approx Var\Bigg(\sum_{j'=1}^{r}\left(\frac{\tau}{nh_0(Y)}f_{jj'}(Y)+\frac{\tau^2}{2n^2h_0^2(Y)}g_{jj'}(Y)\right)\Bigg),\\
&=\sum_{j'=1}^{r}Var\left(\frac{\tau}{nh_0(Y)}f_{jj'}(Y)+\frac{\tau^2}{2n^2h_0^2(Y)}g_{jj'}(Y)\right),\\
&=\sum_{j'=1}^{r}\Bigg(\left(\frac{f_{jj'}(Y)}{nh_0(Y)}\right)^2Var(\tau)+\left(\frac{g_{jj'}(Y)}{2n^2h_0^2(Y)}\right)^2Var(\tau^2)\\
&+\frac{2f_{jj'}(Y)g_{jj'}(Y)}{2n^3h_0^3(Y)}Cov(\tau,\tau^2)\Bigg),\\
&=\sum_{j'=1}^{r}\Bigg(\left(\frac{f_{jj'}(Y)}{nh_0(Y)}\right)^2\frac{k}{h_0^2(Y)}+\left(\frac{g_{jj'}(Y)}{2n^2h_0^2(Y)}\right)^2\frac{2k(k+1)(2k+3)}{h_0^4(Y)}\\
&+\frac{2f_{jj'}(Y)g_{jj'}(Y)}{2n^3h_0^3(Y)}\frac{2k(k+1)}{h_0^3(Y)}\Bigg).
\end{aligned}
\tag{3.70}
$$

Inserting Equation (3.69) into the required leap condition gives the following inequality.

$$
\left|\sum_{j'=1}^{r}\left(\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}k+C(Y)\right)^2-\big(C(Y)\big)^2\right)\right|\leq\epsilon h_0(Y),
\tag{3.71}
$$

where $C(Y)=\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}$. Then, the inequality in (3.71) implies that

$$
\begin{aligned}
\sum_{j'=1}^{r}\big(C(Y)\big)^2-\epsilon h_0(Y)&\leq\sum_{j=1}^{r}\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}k+C(Y)\right)^2\\
&\frac{\sqrt{\sum_{j'=1}^{r}\big(C(Y)\big)^2-\epsilon h_0(Y)}-C(Y)}{\sum_{j'=1}^{r}\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}}\leq k
\end{aligned}
\tag{3.72}
$$

and

$$\sum_{j'=1}^{r}\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}k+C(Y)\right)^2\leq\sum_{j'=1}^{r}\big(C(Y)\big)^2+\epsilon h_0(Y),$$

where $C(Y)=\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}$. Here, the aim of defining $C(Y)$ is just for a simplification of the writing.

$$k\leq\frac{\sqrt{\sum_{j'=1}^{r}\left(\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}\right)^2-\epsilon h_0(Y)}-\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}}{\sum_{j'=1}^{r}\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}}. \tag{3.73}$$

Similarly, $Var(\Delta h_j(Y))$ gives the expression below.

$$k\leq\frac{\sqrt{\epsilon^2h_0^2(Y)+\sum_{j'=1}^{r}\left(\frac{19g_{jj'}^2(Y)}{4n^4h_0^8(Y)}-\frac{8f_{jj'}(Y)g_{jj'}(Y)}{n^3h_0^6(Y)}-\frac{3f_{jj'}^2(Y)}{n^2h_0^4(Y)}\right)}-\left(\frac{5g_{jj'}(Y)}{2n^2h_0^4(Y)}+\frac{2f_{jj'}(Y)}{nh_0^2(Y)}\right)}{\sum_{j'=1}^{r}\frac{g_{jj'}(Y)}{n^2h_0^4(Y)}}. \tag{3.74}$$

Then, from inequalities in (3.73) and (3.74), it can be inferred that

$$k=\min_j\left\lceil\frac{\sqrt{\sum_{j'=1}^{r}\left(\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}\right)^2-\epsilon h_0(Y)}-\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}}{\sum_{j'=1}^{r}\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}}\right.$$
$$\left.\frac{\sqrt{\epsilon^2h_0^2(Y)+\sum_{j'=1}^{r}\left(\frac{19g_{jj'}^2(Y)}{4n^4h_0^8(Y)}-\frac{8f_{jj'}(Y)g_{jj'}(Y)}{n^3h_0^6(Y)}-\frac{3f_{jj'}^2(Y)}{n^2h_0^4(Y)}\right)}-\left(\frac{5g_{jj'}(Y)}{2n^2h_0^4(Y)}+\frac{2f_{jj'}(Y)}{nh_0^2(Y)}\right)}{\sum_{j'=1}^{r}\frac{g_{jj'}(Y)}{n^2h_0^4(Y)}}\right\rceil. \tag{3.75}$$

Thus, as MLE gives the sufficient statistics and as a consequence of the Rao-Blackwell theorem [3], we can theoretically say that Equation (3.75) obtained from (3.75) can provide more accurate results than the estimators derived by the method of moments.

Moreover, the confidence interval for the value of k estimated from MLE is found as previously, $k\approx\frac{k}{nh_0^2(Y)}-\chi^2_{\alpha/2}\sqrt{\frac{k}{h_0^4(Y)n^3}}$ and $k\approx\frac{k}{nh_0^2(Y)}+\chi^2_{1-\alpha/2}\sqrt{\frac{k}{h_0^4(Y)n^3}}$. Then, insterting this interval into Equation (3.42) gives the following approximation.

$$\Delta h_j(Y)\approx\sum_{j'=1}^{r}\Bigg(\left(\frac{k}{nh_0^2(Y)}-\chi^2_{\alpha/2}\sqrt{\frac{k}{h_0^4(Y)n^3}}\right)f_{jj'}(Y)$$
$$+\frac{1}{2}\left(\frac{k}{nh_0^2(Y)}-\chi^2_{\alpha/2}\sqrt{\frac{k}{h_0^4(Y)n^3}}\right)^2g_{jj'}(Y)\Bigg)$$

and

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} \Bigg(\Bigg(\frac{k}{nh_0^2(Y)} + \chi^2_{1-\alpha/2} \sqrt{\frac{k}{h_0^4(Y)n^3}} \Bigg) f_{jj'}(Y) + \frac{1}{2} \Bigg(\frac{k}{nh_0^2(Y)} + \chi^2_{1-\alpha/2} \sqrt{\frac{k}{h_0^4(Y)n^3}} \Bigg)^2 g_{jj'}(Y) \Bigg).$$

After some more steps, $\Delta h_j(Y)$ can be presented as

$$\Delta h_j(Y) = \sum_{j'=1}^{r} \Bigg(\frac{g_{jj'}(Y)}{2n^2 h_0^4(Y)} k^2 - \chi^2_{\alpha/2} \frac{2}{n^2 \sqrt{n} h_0^4(Y)} k\sqrt{k} + \Bigg(\frac{f_{jj'}(Y)}{nh_0^2(Y)} - (\chi^2_{\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^3 h_0^4(Y)} \Bigg) k - \chi^2_{\alpha/2} \frac{f_{jj'}(Y)}{n\sqrt{n} h_0^2(Y)} \sqrt{k} \Bigg) \tag{3.76}$$

and

$$\Delta h_j(Y) = \sum_{j'=1}^{r} \Bigg(\frac{g_{jj'}(Y)}{2n^2 h_0^4(Y)} k^2 + \chi^2_{1-\alpha/2} \frac{2}{n^2 \sqrt{n} h_0^4(Y)} k\sqrt{k} + \Bigg(\frac{f_{jj'}(Y)}{nh_0^2(Y)} + (\chi^2_{1-\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^3 h_0^4(Y)} \Bigg) k + \chi^2_{1-\alpha/2} \frac{f_{jj'}(Y)}{n\sqrt{n} h_0^2(Y)} \sqrt{k} \Bigg), \tag{3.77}$$

where

$$f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i}$$

and

$$g_{jj'}(Y) = \sum_{l=1}^{n} \sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}.$$

As $k = \frac{\tau}{nh_0(Y)}$ estimated from MLE, it is substituted into Equation (3.76) and Equation (3.77). Hence,

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} \Bigg(\frac{g_{jj'}(Y)}{2n^2 h_0^4(Y)} \frac{\tau^2}{n^2 h_0^2(Y)} - \chi^2_{\alpha/2} \frac{2}{n^2 \sqrt{n} h_0^4(Y)} \frac{\tau}{nh_0(Y)} \sqrt{\frac{\tau}{nh_0(Y)}} + \Bigg(\frac{f_{jj'}(Y)}{nh_0^2(Y)} - (\chi^2_{\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^3 h_0^4(Y)} \Bigg) \frac{\tau}{nh_0(Y)} - \chi^2_{\alpha/2} \frac{f_{jj'}(Y)}{n\sqrt{n} h_0^2(Y)} \sqrt{\frac{\tau}{nh_0(Y)}} \Bigg) \tag{3.78}$$

and

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r} \Bigg(\frac{g_{jj'}(Y)}{2n^2 h_0^4(Y)} \frac{\tau^2}{n^2 h_0^2(Y)} + \chi^2_{1-\alpha/2} \frac{2}{n^2 \sqrt{n} h_0^4(Y)} \frac{\tau}{nh_0(Y)} \sqrt{\frac{\tau}{nh_0(Y)}} + \Bigg(\frac{f_{jj'}(Y)}{nh_0^2(Y)} + (\chi^2_{1-\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^3 h_0^4(Y)} \Bigg) \frac{\tau}{nh_0(Y)} + \chi^2_{1-\alpha/2} \frac{f_{jj'}(Y)}{n\sqrt{n} h_0^2(Y)} \sqrt{\frac{\tau}{nh_0(Y)}} \Bigg). \tag{3.79}$$

$E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using $E(\tau\sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{3/2}$, evaluated by a similar process of finding $E(\sqrt{\tau}), Var(\sqrt{\tau}) \approx 0, Var(\tau\sqrt{\tau}) \approx \frac{3k^2+2k}{h_0^3(Y)}, Cov(\tau^2, \tau\sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{7/2} - \frac{k^2+k}{h_0^2(Y)}\left(\frac{k}{h_0(Y)}\right)^{3/2}, Cov(\tau^2, \sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{5/2} - \frac{k^2+k}{h_0^2(Y)}\left(\frac{k}{h_0(Y)}\right)^{1/2}, Cov(\tau\sqrt{\tau}, \tau) \approx 0, Cov(\tau\sqrt{\tau}, \sqrt{\tau}) \approx \frac{k}{h_0^2(Y)}, Cov(\tau^2, \tau) = \frac{2k(k+1)}{h_0^3(Y)}$ and $Cov(\tau, \sqrt{\tau}) \approx 0$. Hereby,

$$
\begin{aligned}
E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \Bigg(& \frac{g_{jj'}(Y)}{2n^4 h_0^6(Y)} \frac{k^2+k}{h_0(Y)} - \chi^2_{\alpha/2} \frac{2}{n^4 h_0^5(Y)\sqrt{h_0(Y)}} \frac{k^{3/2}}{h_0^{3/2}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} - (\chi^2_{\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{k}{h_0(Y)} \\
& - \chi^2_{\alpha/2} \frac{f_{jj'}(Y)}{n^2 h_0^2(Y)\sqrt{h_0(Y)}} \sqrt{\frac{k}{h_0(Y)}} \Bigg)
\end{aligned} \tag{3.80}
$$

and

$$
\begin{aligned}
E(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \Bigg(& \frac{g_{jj'}(Y)}{2n^4 h_0^6(Y)} \frac{k^2+k}{h_0(Y)} + \chi^2_{1-\alpha/2} \frac{2}{n^4 h_0^5(Y)\sqrt{h_0(Y)}} \frac{k^{3/2}}{h_0^{3/2}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + (\chi^2_{1-\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{k}{h_0(Y)} \\
& + \chi^2_{1-\alpha/2} \frac{f_{jj'}(Y)}{n^2 h_0^2(Y)\sqrt{h_0(Y)}} \sqrt{\frac{k}{h_0(Y)}} \Bigg),
\end{aligned} \tag{3.81}
$$

$$
\begin{aligned}
Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \Bigg(& \frac{g^2_{jj'}(Y)}{4n^8 h_0^{12}(Y)} \frac{6k^3+9k^2+6k-1}{h_0^4(Y)} - \chi^2_{\alpha/2} \frac{4(3k^2)+2k}{n^8 h_0^{14}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} - (\chi^2_{\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right)^2 \frac{k}{h_0(Y)} \\
& - \chi^2_{\alpha/2} \frac{2g_{jj'}(Y)}{n^8 h_0^{11}(Y)\sqrt{h_0(Y)}} \left(\left(\frac{k}{h_0(Y)}\right)^{7/2} - \frac{k^2+k}{h_0^2(Y)} \left(\frac{k}{h_0(Y)}\right)^{3/2} \right) \\
& + \frac{g_{jj'}(Y)}{n^4 h_0^6(Y)} \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} - (\chi^2_{\alpha/2})^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{2k(k+1)}{h_0^3(Y)} \\
& - \chi^2_{\alpha/2} \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^6 h_0^8(Y)\sqrt{h_0(Y)}} \left(\left(\frac{k}{h_0(Y)}\right)^{5/2} - \frac{k^2+k}{h_0^2(Y)} \left(\frac{k}{h_0(Y)}\right)^{1/2} \right) \\
& - (\chi^2_{\alpha/2})^2 \frac{4f_{jj'}(Y)}{n^6 h_0^8(Y)} \frac{k}{h_0^2(Y)} \Bigg)
\end{aligned} \tag{3.82}
$$

and

$$
\begin{aligned}
Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r} \Bigg(& \frac{g_{jj'}^2(Y)}{4n^8 h_0^{12}(Y)} \frac{6k^3+9k^2+6k-1}{h_0^4(Y)} + \chi_{1-\alpha/2}^2 \frac{4(3k^2)+2k}{n^8 h_0^{14}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + (\chi_{1-\alpha/2}^2)^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right)^2 \frac{k}{h_0(Y)} \\
& + \chi_{1-\alpha/2}^2 \frac{2g_{jj'}(Y)}{n^8 h_0^{11}(Y)\sqrt{h_0(Y)}} \left((\frac{k}{h_0(Y)})^{7/2} - \frac{k^2+k}{h_0^2(Y)} (\frac{k}{h_0(Y)})^{3/2} \right) \\
& + \frac{g_{jj'}(Y)}{n^4 h_0^6(Y)} \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + (\chi_{1-\alpha/2}^2)^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{2k(k+1)}{h_0^3(Y)} \\
& + \chi_{1-\alpha/2}^2 \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^6 h_0^8(Y)\sqrt{h_0(Y)}} \left((\frac{k}{h_0(Y)})^{5/2} - \frac{k^2+k}{h_0^2(Y)} (\frac{k}{h_0(Y)})^{1/2} \right) \\
& + (\chi_{1-\alpha/2}^2)^2 \frac{4f_{jj'}(Y)}{n^6 h_0^8(Y)} \frac{k}{h_0^2(Y)} \Bigg).
\end{aligned} \tag{3.83}
$$

Like Equation (3.66) and Eqaution (3.67), here, in Equation (3.80), Equation (3.81), Equation (3.82) and Equation (3.83), it is not tractable to get alone k. These equations are consisted of many terms with noninteger powers. For this reason, it makes finding the value of k harder. So finding analytic result can be obtained via numeric analysis.

3.5.2 Extension of Leap Condition via 3^{rd} Order Truncated Taylor Expansion

Similar to using 2^{nd} order truncated Taylor expansion, we apply 3^{rd} order truncated Taylor expansion to approximate $\Delta h_j(Y)$ with the leap condition. Then, we have the following statements.

$$
\begin{aligned}
\Delta h_j(Y) \approx & \sum_{i=1}^{n} \bar{\lambda_i}(Y,\tau) \frac{\partial h_j(Y)}{\partial Y_i} + \frac{1}{2} \sum_{i,l=1}^{n} \bar{\lambda_i}^T(Y,\tau) \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \bar{\lambda_i}(Y,\tau) + \\
& \frac{1}{6} \sum_{i,l,m=1}^{n} \frac{\partial^3 h_j(Y)}{\partial Y_i \partial Y_l \partial Y_m} \bar{\lambda_i}(Y,\tau) \bar{\lambda_l}(Y,\tau) \bar{\lambda_m}(Y,\tau),
\end{aligned} \tag{3.84}
$$

where $\bar{\lambda_i}(Y,\tau) = \sum_{j=1}^{r} k_j v_{ji}$, $\bar{\lambda_l}(Y,\tau) = \sum_{j=1}^{r} k_j v_{jl}$ and $\bar{\lambda_m}(Y,\tau) = \sum_{j=1}^{r} k_j v_{jm}$. Accordingly, we can obtained the subsequent statement.

$$
\Delta h_j(Y) \approx k \sum_{j'=1}^{r} f_{jj'}(Y) + \frac{1}{2} k^2 \sum_{j'=1}^{r} g_{jj'}(Y) + \frac{1}{6} k^3 \sum_{j'=1}^{r} l_{jj'}(Y), \tag{3.85}
$$

where

$$
f_{jj'}(Y) = \sum_{i=1}^{r} \nu_{ij'} \frac{\partial h_j(Y)}{\partial Y_i},
$$

$$
g_{jj'}(Y) = \sum_{l=1}^{n} \sum_{i=1}^{n} \nu_{ij'} \frac{\partial^2 h_j(Y)}{\partial Y_i \partial Y_l} \nu_{j'i}
$$

and

$$l_{jj'}(Y) = \sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\frac{\partial^3 h_j(Y)}{\partial Y_i \partial Y_l \partial Y_m}\nu_{j'i}\nu_{j'l}\nu_{j'm}$$

Similar to previous derivations, under the Gamma distribution, it is known that $k = \tau h_0(Y)$. Subsequently,

$$\Delta h_j(Y) \approx \sum_{j'=1}^{r}\left(\tau h_0(Y) f_{jj'}(Y) + \frac{1}{2}\tau^2 h_0^2(Y) g_{jj'}(Y) + \frac{1}{6}\tau^3 h_0^3(Y) l_{jj'}(Y)\right). \tag{3.86}$$

Then, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are derived by

$$\begin{aligned}
E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r}\Bigg(E(\tau)h_0(Y)f_{jj'}(Y) + \frac{1}{2}E(\tau^2)h_0^2(Y)g_{jj'}(Y) \\
&\quad + \frac{1}{6}E(\tau^3)h_0^3(Y)l_{jj'}(Y)\Bigg), \\
&= \sum_{j'=1}^{r}\Bigg(\frac{k}{h_0(Y)}h_0(Y)f_{jj'}(Y) + \frac{1}{2}\frac{k(k+1)}{h_0^2(Y)}h_0^2(Y)g_{jj'}(Y) \\
&\quad + \frac{1}{6}\frac{k(k+1)(k+2)}{h_0^3(Y)}h_0^3(Y)l_{jj'}(Y)\Bigg), \\
&= \sum_{j'=1}^{r}\left(kf_{jj'}(Y) + \frac{1}{2}k(k+1)g_{jj'}(Y) + \frac{1}{6}k(k+1)(k+2)l_{jj'}(Y)\right)
\end{aligned} \tag{3.87}$$

and

$$\begin{aligned}
Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} Var\left(\tau h_0(Y) f_{jj'}(Y) + \frac{1}{2}\tau^2 h_0^2(Y) g_{jj'}(Y) + \frac{1}{6}\tau^3 h_0^3(Y) l_{jj'}(Y)\right), \\
&= \sum_{j'=1}^{r}\Bigg(h_0^2(Y)f_{jj'}^2(Y)Var(\tau) + \frac{h_0^4(Y)g_{jj'}^2(Y)}{4}Var(\tau^2) \\
&\quad + \frac{h_0^6(Y)l_{jj'}^2(Y)}{36}Var(\tau^3) + h_0^3(Y)f_{jj'}(Y)g_{jj'}(Y)Cov(\tau,\tau^2) \\
&\quad + \frac{h_0^5(Y)g_{jj'}(Y)l_{jj'}(Y)}{6}Cov(\tau^2,\tau^3) + \frac{h_0^4(Y)f_{jj'}(Y)l_{jj'}(Y)}{3}Cov(\tau,\tau^3)\Bigg).
\end{aligned}$$

As $k \sim \Gamma(k, h_(0)(Y))$, it is known that

$$\begin{aligned}
Var(\tau^2) &= \frac{6k^3 + 9k^2 + 6k - 1}{h_0^4(Y)} \\
Var(\tau^3) &= \frac{9k^5 + 72k^4 + 213k^3 + 180k^2 + 210k}{h_0^6(Y)} \\
Cov(\tau,\tau^2) &= \frac{2k(k+1)}{h_0^3(Y)} \\
Cov(\tau^2,\tau^3) &= \frac{k(k+1)(k+2)(6k+12)}{h_0^5(Y)} \\
Cov(\tau,\tau^3) &= \frac{3k(k+1)(k+2)}{h_0^4(Y)}.
\end{aligned}$$

Then,

$$Var(\Delta h_j(Y)) \approx \sum_{j'=1}^{r}\Bigg(f_{jj'}^2(Y)k + \frac{g_{jj'}^2(Y)}{4}(6k^3+9k^2+6k-1) + \frac{l_{jj'}^2(Y)}{36}(9k^5+72k^4+213k^3+180k^2+210k) + f_{jj'}(Y)g_{jj'}(Y)(2k(k+1)) + \frac{g_{jj'}(Y)l_{jj'}(Y)}{6}(k(k+1)(k+2)(6k+12)) + \frac{f_{jj'}(Y)l_{jj'}(Y)}{3}(3k(k+1)(k+2))\Bigg). \tag{3.88}$$

Inserting the statements in Equation (3.87) into the required leap condition results that

$$k \leq \frac{\sqrt{\epsilon h_0(Y) + \sum_{j'=1}^{r}\left(\frac{3(g_{jj'}(Y)+l_{jj'}(Y))^2}{2l_{jj'}(Y)} + f_{jj'}(Y) + \frac{l_{jj'}(Y)}{3}\right)}}{\sum_{j'=1}^{r}\sqrt{\frac{l_{jj'}(Y)}{6}}} - \frac{\left(\frac{g_{jj'}(Y)+l_{jj'}(Y)}{2}\right)\sqrt{\left(\frac{6}{l_{jj'}(Y)}\right)}}{\sum_{j'=1}^{r}\sqrt{\frac{l_{jj'}(Y)}{6}}}. \tag{3.89}$$

As seen above, we can obtain the new values which expected to be more narrower as we expected. Because, using 2^{nd} and 3^{rd} order truncated Taylor expansion are more accurate approximation than using 1^{st} order Taylor expansion.

3.6 Basic Numerical Example

As we mentioned above, the simulation of the result which we have obtained in this study is the one of our future work. However, to show the accuracy of our claim, a basic example is handled in this thesis. Consider the following system.

$$\begin{aligned}
\mathrm{Y}_1 + \mathrm{Y}_2 &\xrightarrow{\mathrm{c}_1} \mathrm{Y}_5\\
\mathrm{Y}_1 &\xrightarrow{\mathrm{c}_2} \mathrm{Y}_1 + \mathrm{Y}_3\\
\mathrm{Y}_5 &\xrightarrow{\mathrm{c}_3} \mathrm{Y}_1 + \mathrm{Y}_2\\
\mathrm{Y}_5 &\xrightarrow{\mathrm{c}_4} \mathrm{Y}_3 + \mathrm{Y}_4\\
2\,\mathrm{Y}_4 &\xrightarrow{\mathrm{c}_5} \mathrm{Y}_1\\
\mathrm{Y}_1 &\xrightarrow{\mathrm{c}_6} 2\,\mathrm{Y}_4\\
\mathrm{Y}_3 &\xrightarrow{\mathrm{c}_7} \emptyset\\
\mathrm{Y}_4 &\xrightarrow{\mathrm{c}_8} \emptyset.
\end{aligned}$$

Hereby, the state vector Y is consisting of $Y_1, \ldots, Y_5$ such that $Y = [Y_1, Y_2, Y_3, Y_4, Y_5]$. Then the vector c which is consisting of the reaction rate constants c_j's for each reaction channel R_j for $j = 1, \ldots, 8$. According the system, stoichoimetry matrix V can be written by $V = \begin{bmatrix} -1 & 0 & 1 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 2 & 0 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

Then, the state $Y = [1, 12, 2, 10, 0]$ and the reaction rate constant $c = [1, 0.5, 1, 0.05, 1, 0.25, 0.1, 0.1]$ are given. Our hazard functions for each reactions are in the following.

$$\begin{aligned} h_1(Y) &= c_1 Y_1 Y_2, \\ h_2(Y) &= c_2 Y_1, \\ h_3(Y) &= c_3 Y_5, \\ h_4(Y) &= c_4 Y_3, \\ h_5(Y) &= c_5 Y_1 (Y_1 - 1)/2, \\ h_6(Y) &= c_6 Y_1, \\ h_7(Y) &= c_7 Y_3, \\ h_8(Y) &= c_8 Y_4. \end{aligned}$$

For this values, $h_0(Y)$ is computed as $h_0(Y) = \sum_{j=1}^{8} h_j(Y) = 59.5$. When substituting this values into Equation (3.7) with $\epsilon = 10^{-2}$, the result is

$$k_{MLE} = \min_j \lceil \{10.3699, 13.3436, 139.4761, \} \rceil. \tag{3.90}$$

So, minimum values of k can be chosen as $k_{MLE} = 10$. To compare the previous result in the literature, we calculate the Equation (2.19). Then, k is chosen from

$$k_{AppGil} = \min_j \lceil \{11.8100, 139.4761\} \rceil. \tag{3.91}$$

Equation (3.91) implies that minimum values of k can be chosen as $k_{AppGil} = 12$.

Thus, Eqaution (3.90) gives the smallest value of k and narrower interval for k when compared to the other. Moreover, if the error control parameter ϵ is chosen larger such as $\epsilon = 10^{-1}$, then the interval is also larger. In other words, the values of k become as follows.

$$k_{MLE} = \min_j \lceil \{113, 123, 13948\} \rceil \tag{3.92}$$

and

$$k_{AppGil} = \min_j \lceil \{118, 13948\} \rceil. \tag{3.93}$$

For this case, MLE also gives the smallest value of k as $k_{MLE} = 113$ and more narrower interval for k. Instead of choosing larger ϵ, when ϵ becomes smaller such as $\epsilon = 10^{-3}$, then

the interval is also smaller.

$$k_{MLE} = \min_j \lceil \{0.7848, 1.6566, 1.3948\} \rceil \tag{3.94}$$

and

$$k_{AppGil} = \min_j \lceil \{1.1810, 1.3948\} \rceil. \tag{3.95}$$

Equation (3.94) and Equation (3.94) gives the same value of k as $k = 1$. However, this is an undesirable situation because it implies the exact solution. We do not want to get such a solution. So we can inferred that error control parameter ϵ should be larger to get a desirable solution for this example. Otherwise, it can become as an exact solution.

This is a basic example of comparison between the result from MLE under poisson distribution and the previous output in the approximate Gillespie algorithm. To obtain more desirable output, a real system and real inputs can be used. The given values are chosen randomly here. However, these arbitrary inputs satify desired results under some cases, not all cases. As a consequence, this example shows that the new value of k is smaller than or equal to the existing value of k in the literature. Furthermore, the new intervals are narrower than the interval which obtained from approximate Gillespie algorithm. In addition to these, the other new outputs in this study can be compared by the similar way.

CHAPTER 4

CONCLUSION

In the biological systems, there are occuring many reactions. These reactions can be simulated with the help of stochastic simulation algorithms such as Gillespie method, the next-reaction method and the first-reaction method. However, they are not computationally useful to simulate the large systems. Because the speed of obtaining the result is slow although SSA gives the exact result. In this way, approximate SSAs such as Poisson τ-leap methos, approximate Gillespie algortihm can be the alternative choice to get the results faster ,i.e., to get less computational demand. Basically, these approximate approaches depend on the leap condition. This condition implies that the chosen time step τ satisfies that there should be no significance change in the propensity function that is the hazard function, during the time interval $[t, t+\tau]$. Moreover, how many times each reaction can be carry out in each small time interval can be computed so that it is possible to move along the systems history axis from one time step to the next, rather than moving from one reaction to the next [2], [11].

In this thesis, considering the leap condition, we have derived the confidence interval for the number of simultaneous reactions k in the time interval. Hereby, we have used maximum likelihood estimator, method of moment estimator and Bayesian estimators to construct more close and more narrower confidence interval for the parameter of k.

By using MLE, under Poisson distribution, we have found Equation (3.4) and it is the same as Equation (2.19) which is obtained from approximate Gillespie algorithm. Also, we have constructed the confidence interval for k at first, then we have inserted it into $\Delta h_j(Y)$ and we have obtained Equation (3.7). That is, k can be chosen among of the following values $\left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} + (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2, \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} - (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2, \frac{\epsilon^2 h_0^2(Y)}{\sum_{j=1}^{r} f_{jj'}^2(Y)}$.

If $\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} > 1$, then we have that

$$\frac{\epsilon^2 h_0^2(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} > \frac{\epsilon^2 h_0^2(Y)}{(\sum_{j=1}^{r} f_{jj'}(Y))^2} > \frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} > \frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}^2(Y)}.$$

Compared with Equation (3.4), it is clear that $\frac{\epsilon^2 h_0^2(Y)}{\sum_{j=1}^{r} f_{jj'}^2(Y)}$ is larger than $\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$. Also,

$\left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} + (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2$ is smaller than $\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$ when $\frac{z^2_{\alpha/2}}{2\sqrt{n}} < \sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} + \left(\frac{z_{\alpha/2}}{2\sqrt{n}}\right)}$. In other words, if $\frac{z^2_{\alpha/2}}{2\sqrt{n}} < \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} - \frac{1}{4}} + \frac{1}{4}\right)$, then we have found more admissable value of k compared with Equation (2.19) and Equation (3.4). In addition to under the Poisson distribution, we have also studied under the Gamma distribution by using MLE. Hereby, in Equation (3.10) the value of $\frac{\epsilon h_0^3(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$ is smaller than the value of $\frac{\epsilon^2 h_0^6(Y)}{\sum_{j=1}^{r} f^2_{jj'}(Y)}$ if $\frac{h_0^3(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} > 1$ as $n \leq 1$. We have that $\frac{\epsilon h_0^3(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} < \frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$ if $h_0(Y) < 1$. With the help of the previous observation, we can say that $\frac{\epsilon h_0^3(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$ is smaller than $\left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} + (\frac{z_{\alpha/2}}{2\sqrt{n}})^2} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2$ for $\frac{z^2_{\alpha/2}}{2\sqrt{n}} \geq \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} - \frac{1}{4}} + \frac{1}{4}\right)$. Under these conditions, we have found the smallest value of k which can generated with more accurate result. Furthermore, Equation (3.17) has acquired by constructing confidence interval for the value of k at the beginning. Then, under the conditions which satisfying $1 < \frac{\epsilon h_0^3(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} < \frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)}$, Equation (3.17) compared with Equation (3.10) implies that Equation (3.17) is smaller than Eqaution (3.10) if $\frac{z^2_{\alpha/2}}{2\sqrt{n}} < \left(\sqrt{\frac{\epsilon h_0(Y)}{\sum_{j=1}^{r} f_{jj'}(Y)} - \frac{1}{4}} + \frac{1}{4}\right)$. However, this condition is a rare as it can be observable when the reaction rate constants c_i's in the calculation of hazard is very close to zero. In most reactions, c_i's are larger. As a result, using MLE under both Poisson and Gamma distribution has given the value of k which generated with more accurate and more narrower result as desired.

As an alterative approach of MLE, we use MME which is one of the well known estimator. Moreover, when compared to the results with MLE and MME, MLE has given more reliable as it gives the optimal solution. Also, the MLE guarantess the condition of the Cramer Rao lower bound, if an efficient estimator for k is found. In other words, it is stable with less variance [1].

Alongside with MLE and MME, by using the study of Sahai and Khurshid (1993) [18], we have found new confidence intervals for the value of k under the Poisson distribution. These values are shown in Equations (3.30), (3.30) and (3.35). Theoretically, we can say that these are also more narrower than Equation (2.19).

Additionally, instead of MLE and MME, we have used the Bayesian approach to obtain the appropriate value of k. As a result, the output has given more flexible alternative confidence intervals for k owing to the controllable significance level α. The value of α is chosen such that satisfying the value of $z = 1$. So, for some results, we use this value to simplfy the expression and to make easier the computation whereas we just leave the statements with general form of z without any specific α.

Lastly, we have expanded the function of the net change in the hazard function with the 2^{nd} and 3^{rd} order truncated Taylor expansion. We have applied the same process as the 1^{st} order Taylor expansion. The results from MME are not tractable because of their high powers, while MLE can be more tractable. In other words, finding a unique solution for the value of

k from MME is not easy analictly. Hereby, it can used optimization or numerical methods to find the unique value of k.

As a consequence, we have used sufficient statistic in the derivation of k. Thus, theoretically, we have found an admissable value of k which can generate closer, more accurate and more narrower results due to Rao-Blackwell theorem [5], [6]. Moreover, from the advantage of MLE, it is known that it gives sufficient statistics, whereas, the MME and Bayesian estimators cannot guarantee sufficient outcomes. But if plausible k is obtained, it can be adjusted by the formula of MME and Bayesian methods or the results of MLE can be a good starting point while obtaining MME and Bayesian estimators to get sufficient statistics. Hence, as a future work, we consider to generate simulation studies with distinct scenarios so that we can numerically evaluate the performance of the found theoretically results.

REFERENCES

[1] L. J. Bain and M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, Brooks/Cole Cengage Learning, 1992.

[2] Y. Cao, D. T. Gillespie, and L. R. Petzold, Avoiding negative populations in explicit poisson tau-leaping, The Journal of Chemical Physics, 123(5), p. 054104, 2005.

[3] G. Casella and R. L.Berger, *Statistical İnference*, Duxbury Thomson Learning, 2002.

[4] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, Binomial distribution based τ-leap accelerated stochastic simulation, The Journal of Chemical Physics, 122(2), p. 024112, 2005.

[5] S. Demirbüken and V. Purutçuoğlu, Extension of leap condition in approximate stochastic simulation algorithms of biological networks, in *Proceeding of the 4th International Conference on Mathematics*, pp. 288–298, 2020.

[6] S. Demirbüken and V. Purutçuoğlu, Extension of leap condition in approximate stochastic simulation algorithms of biological networks with 2^{nd} and 3^{rd} order taylor expansion, in *Conference Proceeding Book of* 2^{nd} *International Conference on Mathematics and Its Applications in Science and Engineering.*, p. in press, 2021.

[7] M. A. Gibson and J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, The Journal of Physical Chemistry A, 104(9), pp. 1876–1889, 2000.

[8] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, The Journal of Physical Chemistry, 81(25), pp. 2340–2361, 1977.

[9] D. T. Gillespie, A rigorous derivation of the chemical master equation, Physica A, 188(404), p. 25, 1992.

[10] D. T. Gillespie, The chemical langevin equation, The Journal of Chemical Physics, 113(1), pp. 297–306, 2000.

[11] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, The Journal of Chemical Physics, 115(4), pp. 1716–1733, 2001.

[12] D. T. Gillespie, Stochastic simulation of chemical kinetics, Annual Review of Physical Chemistry, 58(1), pp. 35–55, 2007, pMID: 17037977.

[13] D. T. Gillespie, Simulation methods in systems biology, pp. 125–167, 2008.

[14] D. T. Gillespie and L. R. Petzold, Improved leap-size selection for accelerated stochastic simulation, The Journal of Chemical Physics, 119(16), pp. 8229–8234, 2003.

[15] Investopedia, What is a monte carlo simulation?, `https://www.investopedia.com/terms/m/montecarlosimulation.asp`, 2021, (Accessed on 16 September 2021).

[16] J. M.Bower and H. Bolouri, *Computational Modelling of Genetic and Biochemical Networks*, MIT press, 2001.

[17] V. Purutçuoğlu and E. Wit, An approximation algorithm based on leap condition for stochastical simulation of biomedical systems ", in *Proceeding of the 4th International Conference "Inverse problems: Modelling and Simulation"*, 2008.

[18] H. Sahai and A. Khurshid, Confidence intervals for the mean of a poisson distribution: a review, Biometrical Journal, 35(7), pp. 857–867, 1993.

[19] T. Tian and K. Burrage, Binomial leap methods for simulating stochastic chemical kinetics, The Journal of Chemical Physics, 121(21), pp. 10356–10364, 2004.

[20] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman and Hall/CRC, 2006.

APPENDIX A

RAO-BLACKWELL THEOREM

Let $X_1, X_2, \ldots, X_n$ have joint pdf $f(x_1, x_2, \ldots, x_n; \theta)$, and let $S = (S_1, S_2, \ldots, S_k)$ be a vector of jointly sufficient statistics for θ. If T is any unbiased estimator of $\tau(\theta)$, and if $T^* = E(T)$, then

1. T^* is an unbiased estimator of $\tau(\theta)$,
2. T^* is a function of S, and
3. Var(T^*) $\leq$ Var(T) for every θ, and Var(T^*)< Var(T) for some θ unless $T^* = T$ with probability 1.

APPENDIX B

DISTRIBUTIONS

B.1 Poisson Distribution

Let $g(\tau) = \sqrt{\tau}$ be a smooth function for $\tau \geq 0$ with $\tau \sim Poi(k)$. Then, by the Taylor expansion around the mean $\mu = E(\tau)$, the following expression can be obtained.

$$g(\tau) = g(\mu) + g'(\mu)(\tau - \mu) + \frac{g''(\mu)(\tau-\mu)^2}{2!} + \frac{g'''(\mu)(\tau-\mu)^3}{3!} + \cdots + \frac{g^t(\mu)(\tau-\mu)^t}{t!} + \ldots$$

Then, the mean can be derived as

$$\begin{aligned} E(g(\tau)) &= g(\mu) + g'(\mu)E(\tau-\mu) + \frac{g''(\mu)E(\tau-\mu)^2}{2!} + \frac{g'''(\mu)E(\tau-\mu)^3}{3!} + \ldots \\ &\quad + \frac{g^t(\mu)E(\tau-\mu)^t}{t!} + \ldots \\ &= g(\mu) + g'(\mu)m_1 + \frac{g''(\mu)m_2}{2!} + \frac{g'''(\mu)m_3}{3!} + \cdots + \frac{g^t(\mu)m_t}{t!} + \ldots, \end{aligned}$$

where m_t is t-th central moment. In this case, considering just up to third order Taylor expansion, $m_1 = 0$ and $m_2 = m_3 = \mu$. So, we have

$$E[g(\tau)] = \sqrt{\mu} + 0 + \frac{1}{8}\mu^{-\frac{1}{2}} - \frac{1}{16}\mu^{-\frac{3}{2}}.$$

Then, $E[g(\tau)] = E[\sqrt{\tau}] \approx \sqrt{\mu} = \sqrt{E(\tau)} = \sqrt{k}$ for $\mu >> 1$. Thus, $E(\sqrt{\tau}) \approx \sqrt{k}$.

B.2 Gamma Distribution

Similarly, we apply these processes for the gamma distribution with $\tau \sim \Gamma(k, h_0(Y))$ and by this way, t-th moment for the gamma distribution is defined as $E(\tau^t) = \frac{(k+t-1)\ldots(1)}{h_0^t(Y)}$. Then, we can obtain $E(\tau)$ as below.

$$E(\tau) = \sqrt{\frac{k}{h_0(Y)}} + \frac{1}{8}\left(\frac{1}{\sqrt{k.h_0(Y)}} - \frac{1}{h_0^4(Y).k\sqrt{k}}\right),$$

with $E[\tau-\mu] = 0$, $E[(\tau-\mu)^2] = V(\tau) = \frac{k}{h_0^2(Y)}$ and $E[(\tau-\mu)^3] = \frac{2k}{h_0^3(Y)}$. If $k \times h_0(Y) << 1$, then it is possible to reach that $E[\sqrt{\tau}] = \sqrt{\frac{k}{h_0(Y)}}$. Similarly, the equality $V(\sqrt{\tau}) = \sqrt{\frac{k}{h_0^2(Y)}}$ can be obtained.