Turk. J. Math. Comput. Sci.
14(2)(2022) 366–375
© MatDer
DOI : 10.47000/tjmcs.901339

Extension of Leap Condition in Approximate Stochastic Simulation Algorithms of Biological Networks

Saliha Demirbüken[1,2], Vilda Purutçuoğlu[3,*]

[1]*Scientific Computing, Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey.*
[2]*Department of Mathematics, Faculty of Science, Gebze Technical University, Kocaeli, Turkey.*
[3]*Department of Statistics, Faculty of Arts and Sciences, Middle East Technical University, Ankara, Turkey.*

Received: 22-03-2021 • Accepted: 05-04-2022

Abstract. In the biological systems, Monte Carlo approaches are used to provide the stochastic simulation of the chemical reactions. The major stochastic simulation algorithms (SSAs) are the direct method, also known as the Gillespie algorithm, the first reaction method and the next reaction method. While these methods give accurate generation of the results, they are computationally demanding for large complex systems. To increase the computational efficiency of SSAs, approximate SSAs can be option. The approximate methods rely on the leap condition. This condition means that the propensity function during the time interval t to $[t + \tau]$ should not be altered for the chosen time step τ. Here, to proceed with the system's history axis from one time step to the next, we compute how many times each reaction can be realized in each small time interval τ so that we can observe plausible simultaneous reactions. Hence, this study aims to generate a realistic and close confidence interval for the parameter which denotes the underlying numbers of simultaneous reactions in the system by satifying the leap condition. For this purpose, the poisson τ-leap algorithm and the approximate Gillespie algorithm, as the extension of the Gillespie algorithm, are handled. In the estimation for the associated parameters in both algorithms, we derive their maximum likelihood estimators, moment estimators and bayesian estimators. From the derivations, we theoretically show that our novel confidence intervals are narrower than the current confidence intervals under the leap condition.

2010 AMS Classification: 92-10, 62E17, 92C42

Keywords: Approximate stochastic simulation algorithms, leap condition, confidence interval.

1. Introduction

In the activation of biological systems, many reactions can happen simultaneously. These reactions can be numerically and exactly simulated in the time evolution by the help of the stochastic simulation algorithms (SSAs) [1, 11–13, 15, 21]. The direct method (i.e., Gillespie algorithm), the first reaction method and the next reaction method are the main methods in this field [10, 13]. Although the SSAs are successful in generating the biological systems, their calculations are computationally inefficient. Therefore, they are not preferable for the simulation of realistically complex systems. Hence, the approximate SSAs are suggested to reduce the computational time by loosing accuracy regarding the exact SSAs. Mainly, the approximate SSAs are based on the *leap condition* [16, 17]. This condition implies that the time step τ should be limited so that there can be no serious difference in the values of the propensity

*Corresponding Author
Email addresses: salihademirbuken@gmail.com (S. Demirbüken), vpurutcu@metu.edu.tr (V. Purutçuoğlu)

function during the time difference between initial time t and time after the time step τ, $[t+\tau]$ [13, 20]. This condition is mathematically shown as below [19].

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

where $h_j(Y)$ is the hazard of jth reaction at the state Y for $j = 1,\ldots,r$, i.e, the propensity function and $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$ is the sum of all hazard functions. Moreover, ϵ stands for the error control parameter. In this expression, j is the indicator of the reaction for a system composed of r reactions, i.e, $j = 1, 2, ..., r$ and $\bar{\lambda}$ is computed as below.

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

Hence, this statement represents that the expected net change in the state Y for the given time interval τ with r numbers of reactions in the system. Also, v_j denotes the stoichiometric coefficients of the reaction j which corresponds to the jth row of the net effect matrix V.

Generally, finding $\hat{Y}$, which is the reasonable value of an unknown population parameter by the observations on having data about the population, can be possible by the estimators. To estimate $\hat{Y}$, different methods are used. In this work, we apply the maximum likelihood estimators (MLE), moment estimator (ME) and bayesian estimators to infer the parameters, namely, the number of simultaneous reactions k, which are occuring at the same time in the same system, and the associated time interval to realize k number of reactions, i.e., τ.

In these calculators, we initially define the following equality;

$$L_n(\theta) = L_n(\theta, y) = f_n(y;\theta),$$

where the observed data set is denoted by $y = (y_1, y_2, \ldots, y_n)$, associated with a vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]^{\mathsf{T}}$ of parameters that index the probability distribution within a parametric family $\{f(\cdot\,;\theta) \mid \theta \in \Theta\}$. Hence, this expression is called the parameter space and $f_n(y;\theta)$ is defined as the product of univariate density functions. Consequently, MLE aims to derive the values of the model parameter that enlarge the likelihood function over the parameter space, i.e,

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \widehat{L}_n(\theta\,; y).$$

On the other hand, in order to define the ME expressions, we think the problem of estimation as the following way: Let k denote the unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$ characterizing the distribution $f_W(w;\theta)$ of the random variable W. Accordingly, considering that the first m moments of the true distribution W_m $(m = 1, 2, \ldots)$ exist, we can denote these moments as the function of θ's by

$\mu_1 \equiv \mathrm{E}[W] = g_1(\theta_1, \theta_2, \ldots, \theta_m)$,

$\mu_2 \equiv \mathrm{E}[W^2] = g_2(\theta_1, \theta_2, \ldots, \theta_k)$,

$\vdots$

$\mu_k \equiv \mathrm{E}[W^k] = g_k(\theta_1, \theta_2, \ldots, \theta_k)$,

in which $E[.]$ presents the expectation of the given function and $g_i(i = 1, \ldots, k)$ shows the functions of θ_i.

Finally, the bayesian estimator can be defined as follows: Let $R_T(\theta)$ be the risk function, i.e., $R_T(\theta) = E[L(T;\theta)]$. Here, $L(T;\theta)$ is a loss function such that $L(t;\theta) \geq 0$ for every t and $L(t;\theta) = 0$ when $t = \tau(T(\theta))$ if T is an estimator of $\tau(\theta)$. Then, the bayesian estimator T^* relative to the risk function $R_T(\theta)$ and the associated distribution function $p(\theta)$ are the estimator with a minimum expected risk via

$$E_\theta[R_{T^*}] \leq E_\theta[R_T]$$

for every estimator T. Here, $p(\theta)$ is usually defined as the prior density for the parameter θ. In other words, $p(\theta)$ represents the prior knowledge or belief about the true value of the parameter. Additionally, the bayesian approach selects a suitable prior belief by comparing the risk functions and then, finds the best estimator of the parameter [2], [6].

In addition to MLE, ME and bayesian estimator as the point estimator, it is also possible to gain an estimated range of values containing the unknown population parameter with the help of the confidence interval.

Hereby, this work aims to generate the confidence interval for the population parameters k and τ by using MLE, ME and bayesian estimators under two approximate SSAs, namely, the poisson τ-leap, which is the most well-known

approximate SSAs, and its extension, called, the approximate Gillespie algorithm. Therefore, we reproduce the leap condition and construct one-sided confidence interval for k and τ under a given significance level α. By this way, we can generate confidence intervals for both parameters by controlling α. Generating confidence intervals based on α and estimating the model parameters via distinct inference approaches are two major outcomes of this study. With the help of these novelties, we can obtain narrow confidence intervals for the plausible values of k and τ. Accordingly, in the organization of the study, we present the general idea of approximate SSAs and the mathematical details of the poisson τ-leap and approximate Gillespie approach in Section 2. We introduce our confidence intervals in Section 3 and in Section 4, we conclude our results.

2. Approximate Stochastic Simulation Algorithms

The stochastic simulation algorithms SSAs are preferable since they can exactly generate the biological systems' activation. However, as they are computationally costly, they are not suggested for large systems. To overcome this problem, the common way is to use the approximate SSAs. In other words, applying the approximate SSAs can gain from the computational demand. In this way, it is possible to progress with one step to the next step, rather than the reaction to the next in the history axis of the system, by controlling how many times each reaction can be realized in each sub-interval, i.e, the leap. When the smaller time interval is selected, then the more accurate results are obtained. However, when the larger time interval is chosen, the approximated results can be obtained with high computational cost [13], [18]. In this part, we initially present the poisson τ-leap method as a common approximate SSA and then, we represent the approximation Gillespie algorithm as the extension of it.

2.1. Poisson τ-leap Method.

In this algorithm, under the leap condition, a random value k_j is generated from a poisson distribution via $Poi(h_j(Y)\tau)$ for each reaction channel in the time interval $[t, t+\tau]$ [13]. Here, $Y(t) = Y$ is a state vector. Then, an admissible τ is found by inserting it into the following inequality;

$$|h_j(Y + \lambda(\bar{Y})) - h_j(Y)|,$$

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

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

which is the expected net change in the state for the given time interval. Here, v_j is the stochiometric coefficients of the reaction j corresponding to the jth row of the net effect matrix V as stated beforehand and $h_j(Y)$ is the hazard function of the jth reaction that is found by the product of the rate constant c_j and the distinct molecular reactant combination of the underlying reaction. Then , $\xi_i(Y) = \sum_{j=1}^{r} h_j v_{ij}$ can be shown as the mean or the expected state change in a unit of time by an n-dimensional vector where each ith component, $\xi_i(Y)$, corresponds to the expected change of the ith species in a unit of time. Then, the following expression is obtained;

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

by using $\bar{\lambda}(Y,\tau)$ in Equation (2.1). It can be inferred that the expected changes in a hazard functions in the time τ are limited by a fraction ϵ, error control parameter lying $(0 < \epsilon < 1)$ and the sum of all hazard functions $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$. Indeed, this inequality (2.2) gives the leap condition. Consequently, after the expansion of the first order Taylor formula and its application on Equation (2.2), the followings equality is found;

$$h_j(Y + \bar{\lambda}(Y,\tau)) \approx \bar{\lambda}(Y,\tau) h_j'(Y) = \sum_{i=1}^{n} \tau\xi_i(Y) \frac{\partial h_j(Y)}{\partial Y_i}.$$

Then, letting $b_{ji}(Y) = \frac{\partial h_j(Y)}{\partial Y_i}$ $(i = 1, ..., n;\ j = 1, ..., r)$, the below inequality can be obtained;

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

As a result, the largest value of τ satisfying the leap condition for the given Y and the pre-selected ϵ is computed by

$$\tau = min\Big\{\frac{\epsilon h_0(Y)}{|\sum_{i=1}^{n} \xi_i(Y) b_{ji}(Y)|}\Big\}. \tag{2.3}$$

It can be accessible that using the exact SSA from Equation (2.3) is more feasible since the reached value of τ is proper for the leap size. The obtained τ in Equation (2.3) would not be chosen if $\tau \leq \frac{1}{h_0(Y)}$ as $\tau = \frac{1}{h_0(Y)}$ is gained from SSA. Regardless of the computational cost, the time interval in the poisson τ-leap method is more preferable than the time of SSA. Actually, there is an incremental difference between them.

The final step is to update the current state in the poisson τ-leap method by replacing t by $t := t + \tau$. Also, for Y, there is a requirement to decide the largest value of τ and to be appropriate with the leap condition.

From the application of this method in various systems, it is seen that the poisson τ-leap approach can bring negative molecular populations in the long-run simulations. In order to solve this problem, some alternative solutions are proposed. The binomial τ-leap is one of the these approaches. This method can unravel the negativity problem, whereas, it is not accurate enough to obtain smooth approximation of exact SSAs [3–5, 18, 21].

2.2. **Approximate Gillespie Algorithm.** There are different alternatives of the poisson τ-leap apporach in the literature. Among them, in this study we use the approximate Gillespie algorithm since it is one of the closer approach of poisson τ-leap in distribution sense. Basically, the approximate Gillespie algorithm [19], which is based on the extension of the exact Gillespie method, states that k numbers of reactions, obtained 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. Hence, we can present $\tau \sim \Gamma(k, h_0(Y))$, where τ denotes 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, we assume that the essential time for every reaction is corresponding 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, we initially identify 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.

$$\Delta h_j(Y) = h_j(Y + \lambda(\bar{Y}, \tau)) - h_j(Y) \approx \lambda(\bar{Y}, \tau) h_j(Y) = \sum_{i=1}^{n} \lambda(\bar{Y}, \tau) \frac{\partial h_{ij}(Y)}{\partial Y_i} \tag{2.4}$$

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

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

In the above expression, k_j shows the number of times of the jth reaction fired in $[t, t + \tau]$ and v_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 represent $\tau \sim \Gamma(k, h_0(Y))$ 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 (2.4), we can get

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

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

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

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

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

where $Var(.)$ denotes the variance of the given random variable. Then, the statistics for $\Delta h_j(Y)$ can be shown by

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}(Y)E(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}(Y) \tag{2.5}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}^2(Y)Var(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}^2(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}^2(Y). \tag{2.6}$$

By substituting Equation (2.5) and (2.6) into the required leap condition, the below expression can be found;

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

and

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

Accordingly, the optimal k is computed from

$$k = \min_{j\in[1,r]}\left\lfloor \left|\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)}\right|, \left|\frac{\epsilon^2 h_0^2(Y)}{\sum_{j'=1}^{r} f_{jj'}^2(Y)}\right| \right\rfloor. \tag{2.7}$$

Indeed, 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 [14]. But in both studies, the confidence intervals are constructed one-sided and by taking a fixed significance level α which roughly sets the tabulated value to 1. Moreover, they produce large intervals which decreases the accuracy of the approximate algorithms. Hereby, the following part introduce our proposal confidence intervals which can generate accurate results regarding previous studies.

3. Confidence Intervals for Leap Condition

In order to improve the performance of the underlying approximate SSAs, we have derived the MLE and ME estimators of the parameters in the underlying approach in the study of Demirbüken and Purutçuoğlu [7]. In this study, we insert these estimators by including their bayesian version in the construction of the confidence intervals and extend the derivations. Hence, under the assumption that $\tau \sim \Gamma(k, h_0(Y))$, the MLE of k is found as $\frac{\tau}{nh_0(Y)}$, where $\tau = \sum_i^n \tau_i$. Then, by plugging it into $\Delta h_j(Y)$, the following is

$$\Delta h_j(Y) = \sum_{j=1} f_{jj'}(Y)\frac{\tau}{nh_0(Y)}.$$

Then, with the knowledge of the value of $z_{\alpha/2} = 1$, where $z_{\alpha/2}$ is in the general formula of the confidence interval, similar to Equation (2.5),

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

Since $\tau \sim \Gamma(k, h_0(Y))$, the mean of the value τ is $E(\tau) = \frac{k}{h_0(Y)}$ and the variance of the value τ is found 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 reached, respectively;

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}(Y)\frac{E(\tau)}{nh_0(Y)} = \frac{k}{nh_0^2(Y)}\sum_{j=1}^{r} f_{jj'}(Y), \tag{3.1}$$

$$Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}^2(Y)\frac{Var(\tau)}{n^2h_0^2(Y)} = \frac{k}{n^2h_0^4(Y)}\sum_{j=1}^{r} f_{jj'}^2(Y). \tag{3.2}$$

After inserting Equation (3.1) and (3.2) into the leap condition in Equation (2.2), a suitable k can be derived as

$$k = \min_{j\in[1,r]} \left\lfloor \left| \frac{\epsilon h_0^3(Y) n}{\sum_{j'=1}^{r} f_{jj'}(Y)} \right|, \left| \frac{\epsilon^2 h_0^6(Y) n^2}{\sum_{j'=1}^{r} f_{jj'}^2(Y)} \right| \right\rfloor. \tag{3.3}$$

It can be seen that the value of $\frac{\epsilon h_0^3(Y)n}{\sum_{j'=1}^{r} f_{jj'}(Y)}$ is smaller than the value of $\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)}$ as presented in Equation (2.7) for $0 < nh_0(Y) < 1$. This implies that k found via Equation (3.3) can produce more accurate result.

Then, we apply this idea by using the method of moment estimators (MME) of the parameters. Accordingly, the estimator 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;

$$\begin{aligned} E(\Delta h_j(Y)) &\approx \sum_{j=1}^{r} f_{jj'}(Y) E(\frac{(n-1)n}{\tau} S^2) \\ &= \sum_{j=1}^{r} f_{jj'}(Y) \frac{(n-1)n}{E(\tau)} S^2 \\ &= \frac{h_0(Y) n(n-1) S^2}{k} \sum_{j=1}^{r} f_{jj'}(Y), \end{aligned} \tag{3.4}$$

$$\begin{aligned} Var(\Delta h_j(Y)) &\approx \sum_{j=1}^{r} f_{jj'}(Y) Var(\frac{(n-1)n}{\tau} S^2) \\ &= \frac{h_0^2(Y) n(n-1) S^2}{k} \sum_{j=1}^{r} f_{jj'}^2(Y). \end{aligned} \tag{3.5}$$

Then, similar to application of MLE, after inserting Equation (3.4) and (3.5) into the required leap condition, a favorable k value can be attained from

$$k \leq \min_{j\in[1,r]} \left\lfloor \left| \frac{n(n-1)S^2 \sum_{j=1}^{r} f_{jj'}(Y)}{\epsilon} \right|, \left| \frac{h_0(Y) n(n-1) S^2 \sum_{j=1}^{r} f_{jj'}^2(Y)}{\epsilon} \right| \right\rfloor.$$

In order to find k, the confidence interval can be constructed. Therefore, the formula $E(k) \pm z_{\alpha/2} \sqrt{\frac{Var(k)}{n}}$ as the well-known representation of the confidence interval, gives that $k \approx \frac{k}{nh_0^2(Y)} \pm \frac{z_{\alpha/2}}{nh_0(Y)} \sqrt{\frac{k}{h_0(Y)}}$. In this inequality, $z_{\frac{\alpha}{2}}$ defines the tabulated normal value for the significance level α. Then, substituting this expression into the Equation (2.4) of $\Delta h_j(Y)$, it follows as

$$\Delta h_j(Y) = \sum_{j'=1}^{r} f_{jj'}(Y) \Big(\frac{k}{nh_0^2(Y)} \pm \frac{z_{\alpha/2}}{nh_0(Y)} \sqrt{\frac{k}{h_0(Y)}} \Big). \tag{3.6}$$

Again embedding $k = \frac{\tau}{nh_0(Y)}$ into Equation (3.6), the followings expressions are obtained;

$$\begin{aligned} \Delta h_j(Y) &\approx \sum_{j'=1}^{r} f_{jj'}(Y) \Big(\frac{\tau}{n^2 h_0^3(Y)} \pm \frac{z_{\alpha/2}}{nh_0(Y)} \sqrt{\frac{\tau}{nh_0^2(Y)}} \Big), \\ \Delta h_j(Y) &\approx \sum_{j'=1}^{r} f_{jj'}(Y) \Big(\frac{\tau}{n^2 h_0^3(Y)} \pm \frac{z_{\alpha/2}}{nh_0^2(Y)} \sqrt{\frac{\tau}{n}} \Big). \end{aligned}$$

For this statement, we derive the values of $E(\sqrt{\tau})$ and $Var(\sqrt{\tau})$ (see Appendix). Then, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using the assumption of $E(\sqrt{\tau}) = \sqrt{\frac{k}{h_0(Y)}}$ and $Var(\sqrt{\tau}) = \sqrt{\frac{k}{h_0^3(Y)}}$ by

$$
\begin{aligned}
E(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}(Y)\Big(\frac{E(\tau)}{n^2h_0^3(Y)} \pm z_{\alpha/2}\frac{E(\sqrt{\tau})}{nh_0^2(Y)\sqrt{n}}\Big) \\
&= \sum_{j'=1}^{r} f_{jj'}(Y)\Big(\frac{k}{n^2h_0^4(Y)} \pm z_{\alpha/2}\frac{\sqrt{k}}{nh_0^2(Y)\sqrt{nh_0(Y)}}\Big) \\
&= \Big(\frac{k}{n^2h_0^4(Y)} \pm z_{\alpha/2}\frac{\sqrt{k}}{nh_0^2(Y)\sqrt{nh_0(Y)}}\Big)\sum_{j'=1}^{r} f_{jj'}(Y),
\end{aligned} \tag{3.7}
$$

$$
\begin{aligned}
Var(\Delta h_j(Y)) &\approx \sum_{j'=1}^{r} f_{jj'}^2(Y)\Big(\frac{Var(\tau)}{n^4h_0^6(Y)} \pm z_{\alpha/2}^2\frac{Var(\sqrt{\tau})}{n^3h_0^4(Y)}\Big) \\
&= \sum_{j'=1}^{r} f_{jj'}^2(Y)\Big(\frac{k}{n^4h_0^8(Y)} \pm z_{\alpha/2}^2\frac{\sqrt{k}}{n^3h_0^6(Y)}\Big) \\
&= \Big(\frac{k}{n^4h_0^8(Y)} \pm z_{\alpha/2}^2\frac{\sqrt{k}}{n^3h_0^6(Y)}\Big)\sum_{j'=1}^{r} f_{jj'}^2(Y).
\end{aligned} \tag{3.8}
$$

Then, by inserting Equation (3.7) into the required leap condition, the following inequality is derived;

$$
\begin{aligned}
\Big|\Big(\frac{k}{n^2h_0^4(Y)} \pm z_{\alpha/2}\frac{\sqrt{k}}{nh_0^2(Y)\sqrt{nh_0(Y)}}\Big)\sum_{j'=1}^{r} f_{jj'}(Y)\Big| &\le \epsilon h_0(Y) \\
\Big|\Big(\frac{k}{n^2h_0^4(Y)} \pm z_{\alpha/2}\frac{\sqrt{k}}{nh_0^2(Y)\sqrt{nh_0(Y)}} + \frac{z_\alpha^2}{4nh_0(Y)} - \frac{z_{\alpha/2}^2}{4nh_0(Y)}\Big)\sum_{j'=1}^{r} f_{jj'}(Y)\Big| &\le \epsilon h_0(Y) \\
\Big|\Big(\Big(\frac{\sqrt{k}}{nh_0^2(Y)} \pm \frac{z_{\alpha/2}}{2\sqrt{nh_0(Y)}}\Big)^2 - \frac{z_{\alpha/2}^2}{4nh_0(Y)}\Big)\sum_{j'=1}^{r} f_{jj'}(Y)\Big| &\le \epsilon h_0(Y) \\
\Big|\Big(\Big(\frac{\sqrt{k}}{nh_0^2(Y)} \pm \frac{z_{\alpha/2}}{2\sqrt{nh_0(Y)}}\Big)^2 - \frac{z_{\alpha/2}^2}{4nh_0(Y)}\Big)\Big| &\le \frac{\epsilon h_0(Y)}{|\sum_{j'=1}^{r} f_{jj'}(Y)|}.
\end{aligned}
$$

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

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

Applying the same process for Equation (3.8), a suitable k can be computed by

$$
\begin{aligned}
k \le \min_{j\in[1,r]} \Bigg\lfloor \Bigg| &\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j=1}^{r} f_{jj'}(Y)|} + \frac{z_{\alpha/2}^2}{4nh_0(Y)}} \mp \frac{z_{\alpha/2}}{2\sqrt{nh_0(Y)}}\right)^2 n^2h_0^4(Y)\Bigg|, \\
&\Bigg|\left(\sqrt{\frac{\epsilon h_0(Y)}{|\sum_{j=1}^{r} f_{jj'}^2(Y)|} + \frac{z_{\alpha/2}^4}{4n^2h_0^4(Y)}} \mp \frac{z_{\alpha/2}^2}{2nh_0^2(Y)}\right)^2 n^4h_0^8(Y)\Bigg|\Bigg\rfloor.
\end{aligned}
$$

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 $\tau \sim Poi(k)$ and $k \sim \Gamma(\alpha,\beta)$, where α and β are the given parameters [2]. 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

$$E(\tau|k) = \frac{\sum \tau_i + \beta}{n + \frac{1}{\alpha}}. \tag{3.9}$$

By inserting Equation (3.9) 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 $\tau \sim Poi(k)$;

$$\begin{aligned} E(\Delta h_j(Y)) &\approx \sum f_{jj'}(Y) \frac{E(\tau) + \beta}{n + \frac{1}{\alpha}} \\ &= \frac{k + \beta}{n + \frac{1}{\alpha}} \sum f_{jj'}(Y), \end{aligned}$$

$$\begin{aligned} Var(\Delta h_j(Y)) &\approx \sum \frac{f^2_{jj'}(Y)}{\left(n + \frac{1}{\alpha}\right)^2} Var(\tau) \\ &= \frac{k}{\left(n + \frac{1}{\alpha}\right)^2} \sum f^2_{jj'}(Y). \end{aligned}$$

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

$$\begin{aligned} k &\leq \frac{(n + \frac{1}{\alpha})\epsilon h_0(Y)}{|\sum f_{jj'}(Y)|} - \beta. \\ k &\leq \frac{(n + \frac{1}{\alpha})^2 \epsilon^2 h_0^2(Y)}{|\sum f^2_{jj'}(Y)|}. \end{aligned}$$

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

$$k \leq \min_{j \in [1,r]} \left\lfloor \frac{(n + \frac{1}{\alpha})\epsilon h_0(Y)}{|\sum f_{jj'}(Y)|} - \beta, \frac{(n + \frac{1}{\alpha})^2 \epsilon^2 h_0^2(Y)}{|\sum f^2_{jj'}(Y)|} \right\rfloor.$$

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

4. Conclusion

It can be possible to simulate the reaction of biological systems. To do this, SSAs play important roles. But, they are not computationally efficient to generate large systems. To reduce computational cost for these approaches, approximate SSAs are the alternative choice. The major aim of these methods depends on the leap condition which implies the time step τ does not alter very much along the time interval $[t, t + \tau]$.

In this work, we have used the properties of the gamma and poisson distribution to obtain distinct estimators as performed in the approximate Gillespie and Poisson τ-leap approach in order to produce confidence intervals for the parameters under a controllable significance level. By this way, we have derived theoretically narrower confidence level which can increase the accuracy of the algorithms without losing computational efficiency. Because we have obtained closed form from each derivation.

In the extension of this study, we aim to show whether any other estimators, such as robust estimators, can produce narrower confidence interval for the parameters. Moreover, we consider to evaluate the performance of our theoretical result in a real system via simulation. In addition to this idea, as a future work, we consider to include the approximate

simulation of the system via leap condition into the modeling part as well. From the study of Gebert et al. [8], a system of differential equation model is used to present the time-series gene expression patterns. The proposal approach uses the system of differential equations in modelling. Hence, if we refer a wider step in the leap condition which implies the change in the approximation of the stochastic steps to the approximation of the deterministic steps, it can be possible to conduct a semi parametric model in place of totally deterministic differential equation models in modeling of gene expression data. Moreover in the study of Weber et al. [23], the system parameters of differential equations are estimated by Chebychevian approximation and generalized semi-infinite optimization approaches which depend on the nonparametric inference. Furthermore, in the study of Uğur et al. [22], the Euler and Runge- Kutta type of approximations are performed to discretize the time so that the differential equation model can be applied to represent the genetic networks whose model parameters are estimated by the least square approximation. In the study of Gebert et al. [9], the suggested model of the same data type is the piecewise linear differential equations whose model parameters are estimate by the discretizated approximation of the least square method. Hence, if we obtain a large step of the leap condition and the investigation of convergence in distribution for the leap condition may help to conduct a semi-parametric method, in place of fully non-parametric approaches, in parameter estimation. Additionally, a long run simulated data via the leap condition can help us to view the deterministic pattern of the system which enables us to observe the system behaviour easily and suggest a more realistic model for genetic networks.

Acknowledgement

For this study, the authors thank to the METU research grant (Project No: 10282) for their support. Also, they would like to express their gratitude to the referees of this journal for their very valuable comments and suggestions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors Contribution Statement

All authors have read and agreed to the published version of the manuscript.

Appendix

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

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

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

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

with $E[\tau-\mu]=0, E[(\tau-\mu)^2]=Var(\tau)=\frac{k}{h_0^2(Y)}$ 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 $Var(\sqrt{\tau})=\sqrt{\frac{k}{h_0^2(Y)}}$ can be obtained.

References

[1] Anderson, D.F., *Incorporating postleap checks in tau-leaping*, The Journal of Chemical Physics, **128**(5)(2008), 054103.

[2] Bain, L.J., Engelhardt, M., Introduction to Probability and Mathematical Statistics, Duxbury Press, 1992.

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

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

[5] Cao, Y., Gillespie, D.T., Petzold, L.R., *Efficient stepsize selection for the tauleaping simulation*, Journal of Chemical Physics, **124**(044109)(2006), 1-144109.

[6] Casella, G., Berger, R.L., Statistical ¨Inference, Duxbury Thomson Learning, 2002.

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

[8] Gebert, J., Latsch, M., Pickl, S.W., Radde, N., Weber et al., *Genetic networks and anticipation of gene expression patterns*, AIP Conference Proceedings, **718**(2004), 474.

[9] Gebert, J., Radde, N., Weber, G.W., *Modeling gene regulatory networks with piecewise linear differential equations*, European Journal of Operational Research, **181**(2006), 1148–1165.

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

[11] Gillespie, D.T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, Journal of Computational Physics, **22**(4)(1976), 403–434.

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

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

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

[15] Gillespie D.T., *Stochastic simulation of chemical kinetics*, Annual Review Physical Chemistry, **58**(2006), 35–55.

[16] Gillespie D.T., *Simulation methods in systems biology*, In Proceedings of the formal methods for the design of computer, communication, and software systems 8th international conference on Formal methods for computational systems biology (SFM'08), Springer-Verlag, Berlin, Heidelberg, (2008), 125–167.

[17] Padgett, J.M., Ilie, S., *An adaptive tau-leaping method for stochastic simulations of reaction-diffusion systems*, AIP Advances, **6**(3)(2016), 035217.

[18] Purutçuoğlu, V., Wit, E., *Exact and approximate stochastic simulations of the MAPK pathway and comparisons of simulations results*, Journal of Integrative Bioinformatics, **3**(2006), 1–13.

[19] Purutçuoğlu, V., Wit, E., *An approximation algorithm based on leap condition for stochastical simulation of biomedical systems*, Proceeding of the 4th International Conference- Inverse Problems: Modelling and Simulation, (2008), 151–152.

[20] Rathinam, M., Petzold, L.R., Cao, Y., Gillespie, D.T., *Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method*, The Journal of Chemical Physics, **119**(24)(2003), 12784–12794.

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

[22] Uğur, Ö., Picki, S.W., Weber, G.-., Wünschiers, R., *Operational research meets biology: an algorithmic approach to analyze genetic networks and biological energy production*, OR Spectrum, (2006), 1–18.

[23] Weber, G.W., Alparslan-Gïk, S.Z., S¨yler, B, *A new mathematical approach in environmental and life sciences: gene-environment networks and their dynamics*, Environmental Model Assessment, **14**(2009), 267–288.