

# Applicability of Importance Sampling to Coupled Molecular Reactions

Werner Sandmann

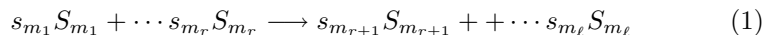
University of Bamberg  
Dep. Information Systems and Applied Computer Science  
Feldkirchenstr. 21  
96045 Bamberg, Germany  
(e-mail: [werner.sandmann@wiai.uni-bamberg.de](mailto:werner.sandmann@wiai.uni-bamberg.de))

**Abstract.** Importance Sampling is a variance reduction technique possessing the potential of zero-variance estimators in its optimal case. It has been successfully applied in a variety of settings ranging from Monte Carlo methods for static models to simulations of complex dynamical systems governed by stochastic processes. We demonstrate the applicability of Importance Sampling to the simulation of coupled molecular reactions constituting biological or genetic networks. This fills a gap between great efforts spent on enhanced trajectory generation and the largely neglected issue of reduced variance among trajectories in the context of biological and genetic networks.

**Keywords:** Stochastic Simulation, Importance Sampling, Molecular Reactions.

## 1 Introduction

As a result of systems' complexity and the huge amount of data that is nowadays available, mathematical modeling and analysis of biological and genetic systems is an emerging area of growing importance. Molecular reactions are the basic building blocks of living systems and essentially biological or genetic networks are formed by coupled molecular reactions. In chemical terminology the fundamental rule of a molecular reaction is given by a stoichiometry



with  $r, \ell \in \mathbb{N}$ ,  $r \leq \ell$ , where  $s_{m_1}, \dots, s_{m_\ell} \in \mathbb{N}$  are *stoichiometric coefficients*,  $S_{m_1}, \dots, S_{m_r}$  are called *reactants*,  $S_{m_{r+1}}, \dots, S_{m_\ell}$  are called *products* and both reactants and products are *molecular species*. Such a chemical equation expresses that the left hand side of the arrow can be transformed to the right hand side of the arrow. Complex chemical processes are given by sets of such reactions. The stoichiometry thus defines which molecular species may react to result in a certain product and how many molecules are involved in a reaction. The temporal behavior is expressed by assigned *reaction rates*. Several mathematical model approaches reflecting different (but related) viewpoints exist for coupled molecular reactions, and the exact meaning of the reaction rates depends on the chosen model type. A comprehensive treatment of modeling approaches can be found in [Bower and Bolouri, 2001].

In the stochastic approach that we adopt in this paper the system state is given by the number of molecules of each species and the transient (time dependent) state probabilities are given by the so-called chemical Master equation. The underlying stochastic process is a Markov jump process and in fact the chemical Master equation is equivalent to the Kolmogorov differential equations. Since direct solution of the chemical Master equation is often analytically intractable, stochastic simulation is in widespread use to analyze systems of coupled molecular reactions, which in its crude version is well-known in the according community as the Gillespie algorithm [Gillespie, 1977]. However, stochastic simulation is inherently costly and besides suffers from the random nature of simulation results. While several attempts to enhance trajectory generation for specifically structured systems have been reported, e.g. [Gillespie, 2001], [Rao and Arkin, 2003], [Rathinam *et al.*, 2003], [Cao *et al.*, 2005], no essential efforts have been spent to reduce the variance among trajectories. We aim at filling this gap by applying Importance Sampling, a well-known classical variance reduction technique, to the simulation of coupled molecular reactions.

The remainder of the paper is organized as follows. Section 2 briefly exposes the stochastic approach to modeling coupled molecular reactions and its relation to Markov processes. In Section 3 the general measure theoretical Importance Sampling setting is given from which the formulae for the application to coupled molecular reactions are derived. Then feasible ways of applying Importance Sampling in this specific setting are investigated. Finally, Section 4 concludes the paper.

## 2 Stochastic Modeling of Coupled Molecular Reactions

Stochastic interpretations of chemically reacting systems can be traced back to the 1960s [McQuarrie, 1967]. A formulation on a physical basis has been provided in [Gillespie, 1976], [Gillespie, 1977] and later on rigorously derived in [Gillespie, 1992]. The basic assumptions are that the system is well stirred and thermally equilibrated, meaning that a well stirred mixture of  $d \in \mathbb{N}^+$  molecular species  $S_1, \dots, S_d$  inside some fixed volume interact at constant temperature. The system state at any time  $t \geq 0$  is a discrete random vector  $X(t) = (X_1(t), \dots, X_d(t))$ , where for each species  $S_k, k \in \{1, \dots, d\}$  and  $t \geq 0$  a discrete random variable  $X_k(t)$  describes the number of molecules of species  $S_k$  present at time  $t$ . The set  $\mathcal{S} \subseteq \mathbb{N}^d$  of all possible system states constitutes the system's state space. The conditional transient (time dependent) probability that the system is in state  $x \in \mathcal{S}$  at time  $t$ , given that the system starts in an initial state  $x_0 \in \mathcal{S}$  at time  $t_0$ , is denoted by

$$p^{(t)}(x) := p^{(t)}(x|x_0, t_0) = P(X(t) = x \mid X(t_0) = x_0). \quad (2)$$

The system state changes due to chemical reactions between molecules of some species. These reactions can be decomposed into unidirectional reaction

channels  $R_1, \dots, R_M$  such that each reaction channel takes the form (1). The reaction rate of each  $R_m, m \in \{1, \dots, M\}$  is given by a well defined function  $\alpha_m$ , called the *propensity function* of reaction channel  $R_m$ , where  $\alpha_m(x)dt$  is the conditional probability that a reaction of type  $R_m$  occurs in the infinitesimal time interval  $[t, t + dt)$ , given that the system is in state  $x$  at time  $t$ . That is  $\alpha_m(x)dt = P(R_m \text{ occurs in } [t, t + dt) \mid X(t) = x)$ . Given that the system starts in an initial state  $x_0 \in \mathcal{S}$  at time  $t_0$ , the temporal evolution of the system is expressed by the *chemical master equation* (CME)

$$\frac{\partial p^{(t)}(x)}{\partial t} = \sum_{m=1}^M \left( \alpha_m(x - v_m) p^{(t)}(x - v_m) - \alpha_m(x) p^{(t)}(x) \right), \quad (3)$$

where  $v_m = (v_{m1}, \dots, v_{md})$  is a *state change vector* and  $v_{mk}, k \in \{1, \dots, d\}$  denotes the change of molecules of species  $S_k$  due to a reaction of type  $R_m$ .

The reaction rates  $\alpha_m$  are time-independent since the probability that a reaction occurs within a specific time interval only depends on the length of this interval and not on the interval endpoints. Thus, given a current system state, the next state in the system's time evolution only depends on this current system state and neither on the specific time nor on the history of reactions that led to the current state. Hence, the time evolution of the system is mathematically described by a stochastic process  $(X(t))_{t \geq 0}$  with  $d$ -dimensional state space  $\mathcal{S} \subseteq \mathbb{N}^d$ , and due to the just stated independence of time and history this stochastic process is a discrete-state Markov process, a Markov jump process, or a continuous-time Markov chain (CTMC).

## 2.1 Equivalence of CME and Kolmogorov Differential Equations

Terminology and notation in the theory of CTMCs is usually rather different from that used to express the CME. Therefore, we briefly explain how they correspond to each other. The multidimensional discrete state space can be mapped to the set  $\mathbb{N}$  of nonnegative integers, i.e. each state  $x \in \mathcal{S}$  is uniquely assigned to an integer  $i \in \{1, \dots, |\mathcal{S}|\}$ . The probability that a transition from state  $i \in \mathbb{N}$  to state  $j \in \mathbb{N}$  occurs within a time interval of length  $h \geq 0$  is denoted by  $p_{ij}(h)$ , and correspondingly  $\mathbf{P}(h) = (p_{ij}(h))_{i,j \in \mathbb{N}}$  is a stochastic matrix, where  $\mathbf{P}(0)$  equals the unit matrix  $\mathbf{I}$ , since no state transitions occur within a time interval of length zero. It is well known (cf. [Bremaud, 1999], [Van Kampen, 1992]) that a CTMC is uniquely defined by an initial probability distribution and a *transition rate matrix*, also referred to as *infinitesimal generator matrix*,  $\mathbf{Q} = (q_{ij})_{i,j \in \mathbb{N}}$  consisting of *transition rates*  $q_{ij}$  where  $\mathbf{Q}$  is the derivative at 0 of the matrix function  $h \mapsto \mathbf{P}(h)$ . The relation of each  $\mathbf{P}(h)$  to  $\mathbf{Q}$  and an explanation for the term *infinitesimal generator matrix* is given by  $\mathbf{P}(h) = \exp(h\mathbf{Q})$ . In that way  $\mathbf{Q}$  generates the the transition probability matrices by a matrix exponential function which is basically defined as an infinite power series. Hence, all information on transition probabilities is covered by the single matrix  $\mathbf{Q}$ . In terms of  $\mathbf{P}$  and

$\mathbf{Q}$  the *Kolmogorov forward differential equations*, the *Kolmogorov backward differential equations*, and the *Kolmogorov global differential equations* can be expressed by (from left to right: forward, backward, global)

$$\frac{\partial}{\partial t}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q}, \quad \frac{\partial}{\partial t}\mathbf{P}(t) = \mathbf{Q}\mathbf{P}(t), \quad \frac{\partial}{\partial t}p^{(t)} = p^{(t)}\mathbf{Q}, \quad (4)$$

where  $p^{(t)}$  denotes the vector of the transient state probabilities corresponding to (2). Explicitly writing the Kolmogorov global differential equations in terms of the coefficients and some algebra yields

$$\frac{\partial p_i^{(t)}}{\partial t} = \sum_{j:j \neq i} p_j^{(t)} q_{ji} - \sum_{j:j \neq i} p_i^{(t)} q_{ij} = \sum_{j:j \neq i} \left( p_j^{(t)} q_{ji} - p_i^{(t)} q_{ij} \right). \quad (5)$$

Now, the equivalence of the CME and the Kolmogorov differential equations can be easily seen by interpreting  $i \in \mathbb{N}$  as the number assigned to state  $x \in \mathcal{S}$ , i.e.  $p_i^{(t)} = p^{(t)}(x)$ ,  $q_{ij} = \alpha_m(x)$  if  $j$  is the number assigned to state  $x + v_m$ , and  $q_{ji} = \alpha_m(x - v_m)$  if  $j$  is the number assigned to state  $x - v_m$ .

## 2.2 Stochastic Simulation

The essential part of any simulation is to imitate the system under consideration. Consequently, simulation of coupled molecular reactions consists of generating trajectories of a CTMC. With the terminology used in the derivation of the CME this is celebrated (though nothing else than a crude direct generation of trajectories, which is known at the latest since the 1950s) as the Gillespie algorithm [Gillespie, 1977] in the biochemical literature:

Init  $t := t_0$  und  $x := x_0$

**repeat**

1. Compute all  $\alpha_m(x)$  and  $\alpha_0(x) := \sum_{m=1}^M \alpha_m(x)$
2. Generate two random numbers  $u_1, u_2$ , uniformly distributed on  $(0, 1)$
3. Generate time  $\tau$  to next reaction:  $\tau = -\ln(u_1)/\alpha_0(x)$
4. Determine reaction type:  $m = \min\{k : \alpha_1(x) + \dots + \alpha_k(x) > u_2 \alpha_0(x)\}$
5. Set  $t := t + \tau$ ;  $x := x + v_m$
6. Store/Collect/Handle Data

**until** "terminating condition"

An equivalent version using a different interpretation of the CTMC dynamics is due to [Gibson and Bruck, 2000].

## 3 Importance Sampling

Importance Sampling is a variance reduction technique that makes use of a change of measure. The original system is simulated under a different

probability measure, and the systematically biased results are weighted by a correcting factor, the likelihood ratio, to yield unbiased estimates. In a general measure theoretic setting, Importance Sampling is based on the Radon-Nikodym theorem, and all applications of Importance Sampling can be derived from this setting. Consider two probability measures  $P$  and  $P^*$  on a measurable space  $(\Omega, \mathcal{A})$ , where  $P$  is absolutely continuous with respect to  $P^*$ , which means that for all  $A \in \mathcal{A}$ ,  $P^*(A) = 0 \Rightarrow P(A) = 0$ . Then, the Radon-Nikodym theorem guarantees that the Radon-Nikodym derivative  $L = dP/dP^*$  exists, and that

$$\forall A \in \mathcal{A}: P(A) = \int_A L(\omega) dP^*. \quad (6)$$

In the context of Importance Sampling the probability measure  $P^*$  is called the *Importance Sampling measure*, and  $L$  is referred to as the *likelihood ratio*. The basic property exploited by Importance Sampling is that expectations with respect to  $P$  are identical to expectations with respect to  $P^*$  when weighting by the likelihood ratio. Let  $L$  be a version of the likelihood ratio and  $Y$  a random variable on  $(\Omega, \mathcal{A})$ . Then

$$E_P[Y] = \int Y(\omega) dP = \int Y(\omega) L(\omega) dP^* = E_{P^*}[YL]. \quad (7)$$

Using a different density or probability distribution/measure is called a *change of measure*, and it is the essential part and the art of Importance Sampling to perform this change of measure such that more accurate estimates can be achieved. Many early applications of Importance Sampling can be found in [Hammersley and Handscomb, 1964]. The framework for stochastic processes, which is of special interest in our setting of coupled molecular reactions has been given in [Glynn and Iglehart, 1989].

### 3.1 Application to Coupled Molecular Reactions

Applying Importance Sampling to coupled molecular reactions first of all requires the distribution or density, respectively, of reaction paths. The discrete state of the system changes due to molecular reactions. Let  $t_1 < t_2 < \dots$  denote the successive time instants at which reactions occur, and  $R_{m_i}$  the reaction type that occurs at time  $t_i$ , where  $m_i \in \{1, \dots, M\}$ . Define  $\tau_i := t_{i+1} - t_i$  the time between the  $i$ -th and the  $(i+1)$ -th reaction. Hence, state  $x(t_i)$  is reached due to the  $i$ -th reaction  $R_{m_i}$  at time  $t_i$  and remains unchanged for a sojourn time of  $\tau_i$  after which the  $(i+1)$ -th reaction  $R_{m_{i+1}}$  occurs at time  $t_{i+1}$  and changes the state to  $x(t_{i+1})$ . Hence, the time evolution of the system is completely described by the sequence of states and corresponding sojourn times, and in compact form  $(x(t_0), \tau_0), (x(t_1), \tau_1), (x(t_2), \tau_2), \dots)$  describes a trajectory. For a trajectory up to the  $R$ -th reaction, considering the Markovian property implying exponentially distributed sojourn times, the reaction

path density is given by

$$p^{(t_0)}(x_0) \cdot \prod_{i=1}^R \alpha_{m_{i-1}}(x(t_{i-1})) \exp(\alpha_0(x(t_{i-1}))\tau_{i-1}), \quad (8)$$

where  $\alpha_0(x(t_{i-1})) := \alpha_1(x(t_{i-1})) + \dots + \alpha_M(x(t_{i-1}))$  just as in the Gillespie algorithm.

Now, in order to perform an Importance Sampling simulation, we need to change the underlying probability measure, which is in the case of coupled molecular reactions determined by the propensity functions. The requirement of absolute continuity leaves us a great freedom in how to change the measure. It is only necessary that all reaction paths that are possible (have positive probability) under the original measure remain possible with Importance Sampling. That means each measure on the sample path space that meets the aforementioned can be considered, even non-Markovian models are allowed as long as they assign positive probabilities to all possible reaction paths. Nevertheless, we should avoid a large increase in trajectory generation efforts compared to the original measure. Thus, obviously the most natural change of measure is to remain in the Markovian world and the easiest way is to simply change the original propensity functions to "Importance Sampling propensity functions"  $\alpha_m^*$  such that for all  $m \in \{1, \dots, M\}$   $\alpha_m^*(x) = 0 \Rightarrow \alpha_m(x) = 0$ ,  $x \in \mathcal{S}$ , or equivalently, starting with the original propensity functions,  $\alpha_m(x) > 0 \Rightarrow \alpha_m^*(x) > 0$ ,  $x \in \mathcal{S}$ . Importance Sampling then generates trajectories according to the changed propensity functions and multiplies the results with the likelihood ratio to get unbiased estimates for the original system. Trajectory generation is thereby performed as before, e.g. by the Gillespie algorithm, where now the changed propensity functions are used, yielding a sequence of states with according sojourn times and reaction path density as in (8). Thus the likelihood ratio becomes

$$L(\omega) = \frac{\prod_{i=1}^R \alpha_{m_{i-1}}(x(t_{i-1})) \exp(\alpha_0(x(t_{i-1}))\tau_{i-1})}{\prod_{i=1}^R \alpha_{m_{i-1}}^*(x(t_{i-1})) \exp(\alpha_0^*(x(t_{i-1}))\tau_{i-1})}, \quad (9)$$

where we have kept the initial distribution unchanged. Rewriting this likelihood ratio yields

$$L(\omega) = \prod_{i=1}^R \frac{\alpha_{m_{i-1}}(x(t_{i-1})) \exp(\alpha_0(x(t_{i-1}))\tau_{i-1})}{\alpha_{m_{i-1}}^*(x(t_{i-1})) \exp(\alpha_0^*(x(t_{i-1}))\tau_{i-1})}, \quad (10)$$

which shows that the likelihood ratio can be efficiently computed in course of the trajectory generation without much extra computational effort by successively updating its value after each reaction. In particular, the unbiased number of molecules can be obtained at any time.

Although naturally arising the change of measure as described above may be too restrictive. In cases where more flexibility is needed, it is possible

to use a different change of measure in each simulation step or propensity functions that depend on the number of already occurred reactions (corresponding to a nonhomogeneous model) or the history of the just executed simulation steps. Formally, define functions  $\beta_m^{(r)}(x(t_0), \dots, x(t_r))$ , where for all  $m \in \{1, \dots, M\} : \alpha_m(x(t_r)) > 0 \Rightarrow \beta_m^{(r)}(x(t_0), \dots, x(t_r)) > 0$ . Then the reaction path density under Importance Sampling is

$$p^{(t_0)}(x_0) \cdot \prod_{i=1}^R \beta_{m_{i-1}}^{(i-1)}(x(t_0), \dots, x(t_{i-1})) \exp(\beta_0(x(t_0), \dots, x(t_{i-1}))\tau_{i-1}) \quad (11)$$

and the corresponding likelihood ratio (leaving the initial distribution unchanged) becomes

$$L(\omega) = \prod_{i=1}^R \frac{\alpha_{m_{i-1}}(x(t_{i-1})) \exp(\alpha_0(x(t_{i-1}))\tau_{i-1})}{\beta_{m_{i-1}}^{(i-1)}(x(t_0), \dots, x(t_{i-1})) \exp(\beta_0(x(t_0), \dots, x(t_{i-1}))\tau_{i-1})}, \quad (12)$$

which can be easily updated after each reaction in course of the simulation.

### 3.2 Further Issues

Now we are done with demonstrating the applicability of Importance Sampling to coupled molecular reactions in that we have given a framework and general rules. An issue that remains open is to concretize these change of measure rules, i.e. how to change the propensity functions in order to achieve variance reduction in practice. The functions  $\alpha_m^*$  or  $\beta_m^{(r)}$ , respectively, must be chosen dependent on the specific model under consideration. In fact, this is an art of Importance Sampling, and a large body of literature exists on change of measure guidelines for specific model classes, e.g. in the context of rare event simulation, a review of which is far beyond the scope of the present paper. The reader is referred to, e.g. [Heidelberger, 1995], [Bucklew, 2004], [Sandmann, 2007].

Stiff systems are of particular interest and difficulty in analyzing coupled molecular reactions. In stiff systems reaction rates differ in orders of magnitude, which arises because reactions occur on multiple time scales meaning that some reactions are much slower than others and occur significantly rarer. Thus, even the generation of one single trajectory becomes very computer time demanding. In this setting Importance Sampling can result in both accelerated trajectory generation and reduced variance among the trajectories.

## 4 Conclusions

We have shown how to apply Importance Sampling to stochastic simulations of coupled molecular reactions. General conditions and different feasible ways

to perform the change of measure have been given, all of which render efficient computation of the involved likelihood ratios possible. Further research includes the study of specific change of measure strategies and its application to a variety of models. In particular, excessive case studies are required to demonstrate the efficiency gains achieved by Importance Sampling.

## References

- [Bower and Bolouri, 2001]J.M. Bower and H. Bolouri, editors. *Computational Modeling of Genetic and Biochemical Networks*. MIT Press, Cambridge, MA, 2001.
- [Bremaud, 1999]P. Bremaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York, 1999.
- [Bucklew, 2004]J.A. Bucklew. *Introduction to Rare Event Simulation*. Springer, New York, 2004.
- [Cao *et al.*, 2005]Y. Cao, –, and et al. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *Journal of Computational Physics*, 206:395–411, 2005.
- [Gibson and Bruck, 2000]M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry, A* 104:1876–1889, 2000.
- [Gillespie, 1976]D.T. Gillespie. A general method for numerically simulating the time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.
- [Gillespie, 1977]D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 71(25):2340–2361, 1977.
- [Gillespie, 1992]D.T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.
- [Gillespie, 2001]D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115:1716–1732, 2001.
- [Glynn and Iglehart, 1989]P.W. Glynn and D.L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35:1367–1392, 1989.
- [Hammersley and Handscomb, 1964]J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Methuen, London, 1964.
- [Heidelberger, 1995]P. Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation*, 5(1):43–85, 1995.
- [McQuarrie, 1967]D.A. McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4:413–478, 1967.
- [Rao and Arkin, 2003]C.V. Rao and A.P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *Journal of Chemical Physics*, 118:4999–5010, 2003.
- [Rathinam *et al.*, 2003]M. Rathinam, –, and et al. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *Journal of Chemical Physics*, 119:12784–12794, 2003.
- [Sandmann, 2007]W. Sandmann. Efficiency of importance sampling estimators. *Journal of Simulation*, 2007. To appear.
- [Van Kampen, 1992]N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, North-Holland, 1992.