Variance bounds for a class of biochemical reactions

Giovanni Pugliese Carratelli * **Ioannis Lestas** *

* *University of Cambridge, Department of Engineering, CB2 1PZ, United Kingdom (e-mail: gp459,icl20@cam.ac.uk)*

Abstract: We consider the problem of quantifying the variance in molecular numbers in biochemical reactions with nonlinear reaction rates. We address this problem for a specific configuration where a spontaneously formed species determines the rate of formation of another species via a nonlinear reaction rate, with the aim being to quantify the variance of this species. By making use of an appropriate decomposition based on Newton series expansion we derive an analytical expression that provides a hard bound for the variance. The bound becomes an equality when the propensities are linear. Furthermore, numerical investigations demonstrate that this is very close to the actual variance also in regimes where the rate of formation of the species is nonlinear.

Keywords: Mathematical biology in general, Kinetics in biochemical problems, Stochastic systems

1. INTRODUCTION

Biochemical systems are often modelled by means of Chemical Master Equations (CMEs) whose moments can be derived analytically only for a few special cases. Numerically evaluating such moments via simulations requires considerable computational effort by generating trajectories through the *Gillespie Algorithm*, (Gillespie (2007)). Alternative approaches that lead to an approximate analysis include Linear Noise Approximations (LNAs), or *Van Kampen's* Ω-Expansion, which involve a linearisation of the reaction rates.

Recently, techniques aimed to compute bounds upon the moments of interest, such as mean or variance, have been investigated. One approach computes bounds on the moments of Markov processes by solving corresponding linear programs, Kuntz et al. (2019). Other approaches rely upon solving steady state moment equations in conjunction with constraints associated with the positive semi-definiteness of moment matrices, see Ghusinga et al. (2017).

In this abstract we present an alternative approach whereby an appropriately formulated expansion based on *Newton* series, is used to derive a hard bound for the variance for a class of biochemical reactions. The bound becomes exact when the reaction rates are linear. Furthermore, unlike LNAs, it is very close to the true variance also when the reaction rates are highly nonlinear.

The abstract is structured as follows. In Section 2 we introduce the notation and the problem under consideration. The main results are given in Section 3. Section 4 provides a numerical evaluation of the results and conclusions are given Section 5. Sketches of the proofs are provided in the appendix and the full derivations will be provided in an extended version of this work.

2. PROBLEM FORMULATION AND NOTATION

2.1 Notation

The following list introduces various symbols used in the text

$\mathbb{E}[X]$	Expectation of random variable X
$\mathbb{R}[A]$	Set of polynomials in A with real valued coefficients
$\mathbb{R}^{>}$	Set of positive real numbers $\{x \in \mathbb{R} : x > 0\}$
$\mathbb{R}^{\geq}$	Set of non negative real numbers $\{x \in \mathbb{R} : x \geq 0\}$
$\mathbb{Z}^{\geq}$	Set of non negative integers $\{0, 1, 2, 3 \ldots\}$
$A \sim \mathcal{P}(x)$	Random variable A has Poisson probability distribution with mean x
$var(X)$	Variance of random variable X
$(A)_k$	Falling factorial of $A \in \mathbb{Z}^{\geq}$: $(A)_k = \prod_{n=0}^{k-1}(A-n)$
$\Delta_k[\cdot]$	k step forward difference, $\Delta_k[f](A) \equiv f(A+k) - f(A)$
$\Delta_k^p[\cdot]$	k step forward difference operator applied p times
$\text{cov}(X, Y)$	Covariance of random variables X and Y

2.2 Problem Formulation

We aim to analyse the following system

$$\begin{aligned} A &\xrightarrow{F} A+1 \\ A &\xrightarrow{\gamma_A A} A-1 \\ B &\xrightarrow{R(A)} B+1 \\ B &\xrightarrow{\gamma_B B} B-1 \end{aligned} \tag{1}$$

where A and B are two considered biochemical species. Random variable $A(t)$ denotes the number of molecules of species A at time t and similarly random variable $B(t)$ denotes the number of molecules of species B at time t.

The parameter $F \in \mathbb{R}^{>}$ is a constant that denotes the rate at which the species A is produced, $R(A) : \mathbb{Z}^{\geq} \to \mathbb{R}^{\geq}$ is the rate of production of the species B and constants $\gamma_A \in \mathbb{R}^{>}$ and $\gamma_B \in \mathbb{R}^{>}$ represent the death rate of each molecule of species A and B respectively.

For any $a, b \in \mathbb{Z}^{\geq}$, we denote by $\mathbb{P}(a, b, t)$ the probability $A(t) = a$ and $B(t) = b$.

The CME, a version of the Chapman Kolmogorov equation for Markov processes, for system (1) is

$$\begin{aligned}\frac{\partial \mathbb{P}(a,b,t)}{\partial t} &= \gamma_A[(a+1)\mathbb{P}(a+1,b,t) - a\mathbb{P}(a,b,t)] \\ &+F[\mathbb{P}(a-1,b,t) - \mathbb{P}(a,b,t)] \\ &+\gamma_B[(b+1)\mathbb{P}(a,b+1,t) - b\mathbb{P}(a,b,t))] \\ &+R(a)[\mathbb{P}(a,b-1,t) - \mathbb{P}(a,b,t)]\end{aligned} \tag{2}$$

The problem we are trying to address is to find an expression of the mean and variance of B. This is a non-trivial problem due to the nonlinear rate $R(A)$.

In this abstract we present a bound on the variance of B obtained by exploiting a discrete expansion of $R(A)$ based on *Newton* series. The bound becomes exact when $R(A)$ is linear. Furthermore, we show via numerical examples that it is very close to the true variance even in regimes where $R(A)$ is highly nonlinear, in contrast to a LNA approximation for the variance.

It should be noted that filtering problems associated with (1) have also been studied in the literature. The optimal causal filter for estimating A when B is observed was derived in Snyder (1975) and the average squared difference between A and B was quantified in Hinczewski and Thirumalai (2014).

The results that will be presented are associated with the equilibrium distribution of species A and B and we therefore make the following assumption

Assumption 1. System (2) reaches a stationary distribution denoted by $\mathbb{P}(a, b)$ *i.e.* $\lim_{t\to+\infty} \mathbb{P}(a, b, t) = \mathbb{P}(a, b)$. Note that from the CME (2) $\mathbb{P}(a, b)$ satisfies the following

$$\begin{aligned}&\gamma_A[(a+1)\mathbb{P}(a+1,a,t) - a\mathbb{P}(a,b,t)] \\ &+F[\mathbb{P}(a-1,a,t) - \mathbb{P}(a,b,t)] \\ &+\gamma_B[(b+1)\mathbb{P}(a,b+1,t) - b\mathbb{P}(a,b,t))] \\ &+R(a)[\mathbb{P}(a,b-1,t) - \mathbb{P}(a,b,t)] = 0\end{aligned} \tag{3}$$

For convenience in the notation we denote by A the random variable representing the number of molecules of species A at equilibrium. Similarly, we denote by B the random variable representing the number of molecules of species B at equilibrium.

It should also be noted that $A \sim \mathcal{P}(F/\gamma_A)$, *i.e.* random variable A has a Poisson distribution with mean F/γ_A which is a known result that follows from the fact that F and γ_A are constant (see *e.g.* van Kampen (2007)).

3. MAIN RESULTS

Before we present our main results we state a condition upon $R(A)$ as Assumption 2.

Assumption 2. $R : \mathbb{Z}^{\geq} \to \mathbb{R}^{\geq}$ is a ratio of polynomials, *i.e.* $R(A) = \{\frac{f}{g} | f, g \in \mathbb{R}[A], g \neq 0 \ \forall A \in \mathbb{Z}^{\geq}\}$.

Note that this assumption is mild within a biological perspective as reaction rates are typically rational functions of molecule numbers.

In order to give our first result we define the quantities in (4) which are needed for the variance bound derived.

$$\sigma_0 = \mathbb{E}_{\mathcal{P}}[R(A)], \quad \sigma_1 = -\mathbb{E}_{\mathcal{P}}[R(A)] + \frac{\mathbb{E}_{\mathcal{P}}[AR(A)]}{\mathbb{E}_{\mathcal{P}}[R(A)]} \tag{4}$$

Note that in (4) $\mathbb{E}_{\mathcal{P}}[\cdot]$ denotes the expectation with respect to the Poisson distribution of the random variable A.

We now give our main result which is an expression that provides a lower bound on the variance of B.

Proposition 3. Consider the system in (3) with $R(A)$ satisfying Assumption 2. The following inequality holds

$$var(B) \geq \frac{\sigma_0(\gamma_A + \gamma_B) + \sigma_1^2 \mathbb{E}_{\mathcal{P}}[A]}{(\gamma_A + \gamma_B)\gamma_B} \tag{5}$$

where σ_0 and σ_1 are given in (4).

Remark 1. If $R(A)$ is a linear function then, then equation (5) holds with equality. This is stated in Proposition 4.

Remark 2. Parameters σ_0 and σ_1 exist and are bounded as stated in Lemma 6 in the appendix. Note that $\mathbb{E}_{\mathcal{P}}[\cdot]$ is the expectation with respect to a Poisson distribution with a known mean and may be computed to an arbitrary high precision by taking a sufficient amount of terms in the evaluation of the expectation.

Below we show that the bound is exact if all transition rates are linear.

Proposition 4. Consider the system in (3). If $R(A) = R_c A$ with $R_c \in \mathbb{R}^{>}$ the bound for the variance in Proposition 3 holds with equality and is given by (6)

$$var(B) = \frac{R_c \mathbb{E}_{\mathcal{P}}[A](R_c + (\gamma_A + \gamma_B))}{\gamma_B(\gamma_A + \gamma_B)} \tag{6}$$

Remark 3. The variance in the case of Proposition 4 can be computed analytically using LNA approaches, see *e.g.* van Kampen (2007) and Lestas et al. (2008).

Finally we give a side result on the covariance of A and B.

Proposition 5. Consider the systems described in (3). Then $cov(A, B) = cov(B, A)$ satisfies

$$cov(A, B) = \sigma_1 \frac{\mathbb{E}_{\mathcal{P}}[A]}{\gamma_A + \gamma_B} \tag{7}$$

where σ_1 is given in (4).

Remark 4. Note that Proposition 5 holds with equality also when $R(A)$ is nonlinear.

The results stated above have a number of useful properties relative to other more conventional approaches for quantifying the variance. As mentioned in Section 1, Proposition 3 provides a bound for the variance when transition rate $R(A)$ is nonlinear whereas existing methods like LNA and numerical simulations give only approximate values. As it will be seen in Section 4 the variance calculation through the LNA is not as close to the variance of B as is the bound provided in Proposition 3. This

is especially important in the case of highly nonlinear propensity functions where LNAs become less accurate.

4. EXAMPLE

In order to evaluate the conservativeness of the bound for the variance given in (5) we numerically investigate how close this is to the actual variance. We select as function $R(A)$ a common non linear function found in biological systems, *i.e.* a *Hill* function $R(A) = \frac{(A/A_0)^{n_h}}{1+(A/A_0)^{n_h}}$ where $n_h \in \mathbb{R}^{>}$ and $A_0 \in \mathbb{R}^{>}$. We set as parameters $n_h = 9$ and $A_0 = 100$. This function clearly fulfills the conditions in Assumption 2.

Fig. 1 shows that the bound is not conservative and that it well represents the behavior of $var(B)$ for different values of $\mathbb{E}_{\mathcal{P}}[A]$. In particular, the error between the bound and the variance is under 1% for each of the considered values of $\mathbb{E}_{\mathcal{P}}[A]$.

The figure also shows a comparison of the bound with the variance of B computed through the LNA of system (3). Note that the computation carried out via LNA performs poorly for certain regimes. Specifically, the performance is particularly poor when A is with high probability in the non linear regime of the Hill function, while it is more accurate when A is predominantly in the linear regime.

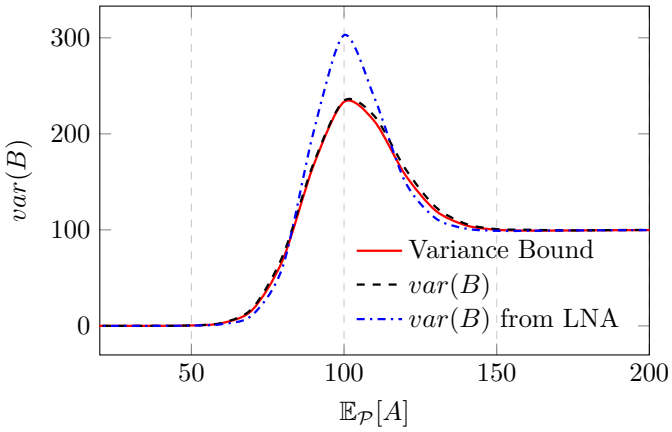

Fig. 1. Numerical simulation showing the bound obtained from (5) and the computed variance of B for varying F/γ_A. $R(A)$ is a Hill function with parameters $n_h = 9$ and $A_0 = 100$. (——) displays the bound computed from (5) and (- - -) displays the variance obtained from numerical simulations of (3). The variance computed through a LNAof system (3) is displayed as (-·-·-)

5. CONCLUSIONS AND FUTURE WORK

We have derived a hard bound for the variance in the molecule numbers of a species formed via another species with a nonlinear reaction rate. The bound follows from an appropriately formulated expansion based on Newton series. This bound holds with equality if the propensity functions are linear and numerical simulations show that this is very close to the actual variance also when the rate of formation of the species is nonlinear. Future work will focus on extending the results to larger classes of CMEs.

REFERENCES

Ghusinga, K.R., Vargas-Garcia, C.A., Lamperski, A., and Singh, A. (2017). Exact lower and upper bounds on stationary moments in stochastic biochemical systems. *Physical Biology*, 14(4), 04LT01. doi:10.1088/1478-3975/aa75c6.

Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1), 35–55. doi:10.1146/annurev.physchem.58.032806.104637.

Hinczewski, M. and Thirumalai, D. (2014). Cellular signaling networks function as generalized wiener-kolmogorov filters to suppress noise. *Physical Review X*, 4(4). doi:10.1103/physrevx.4.041017.

Kuntz, J., Thomas, P., Stan, G.B., and Barahona, M. (2019). Bounding the stationary distributions of the chemical master equation via mathematical programming. *The Journal of Chemical Physics*, 151(3), 034109. doi:10.1063/1.5100670.

Lestas, I., Paulsson, J., Ross, N.E., and Vinnicombe, G. (2008). Noise in gene regulatory networks. *IEEE Transactions on Automatic Control*, 53(Special Issue), 189–200. doi:10.1109/tac.2007.911347.

Snyder, D.L. (1975). *Random Point Processes*. John Wiley & Sons Inc.

van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry*. Elsevier LTD, Oxford.

Appendix A. PROOFS

In the appendix we provide sketches of the proofs of the results presented in the main text. The full derivations will be provided in an extended version of this work.

Lemma 6. Let $A \sim \mathcal{P}(F/\gamma_A)$. Then the quantities σ_0 and σ_1 from (4) exist and are finite.

Proof. The proof follows by noting that σ_0 and σ_1 are expectations of rational functions with respect to a Poisson distribution.

A.1 Sketch proof of Proposition 3

Proof. Using the CME the following expressions for moments of B can be derived

$$\mathbb{E}[B] = \gamma_B^{-1}\mathbb{E}_{\mathcal{P}}[R(A)] \tag{A.1}$$

$$\mathbb{E}[B^2] - \mathbb{E}[B] = \gamma_B^{-1}\mathbb{E}_{\mathcal{P}}[BR(A)] \tag{A.2}$$

Through various algebraic manipulations we express the expectation on the right hand side of (A.1) and (A.2) in terms of coefficients σ_n that can be formulated from the Newton expansion of $R(A)$. In particular, the following polynomial decomposition of $R(A)$ holds (Hinczewski and Thirumalai (2014))

$$R(A) = \sum_{n=0}^{\infty} \sigma_n \psi_n(A) \tag{A.3}$$

where

$$\psi_n(A) = \sum_{k=0}^{n} \binom{n}{k} (-r)^k (A)_{n-k} \tag{A.4}$$

and

$$\sigma_n = \sum_{k=0}^{\infty} \binom{k}{n} \frac{\Delta_1^k[R](0)}{k!} (r)^{k-n} \tag{A.5}$$

where $r \in \mathbb{R}$ is an arbitrary constant. For our derivations we set $r = \mathbb{E}_{\mathcal{P}}[A]$.

By expressing the variance of B in terms of (A.1) and (A.2) an expression for $var(B)$ is obtained as the sum of positive terms

$$var(B) = \gamma_B^{-1} \left[\sigma_0 + \sum_{n=1}^{\infty} \frac{\sigma_n^2 n! (\mathbb{E}_{\mathcal{P}}[A])^n}{\gamma_A n + \gamma_B} \right] \tag{A.6}$$

Since all terms in (A.6) are positive, using only the first two terms we obtain the inequality

$$var(B) \geq \frac{\sigma_0(\gamma_A + \gamma_B) + \sigma_1^2 \mathbb{E}_{\mathcal{P}}[A]}{(\gamma_A + \gamma_B)\gamma_B} \tag{A.7}$$

■

A.2 Sketch proof of Proposition 4

Proof.

When $R(A) = R_c A$ it is possible to show that

$$\begin{aligned} \sigma_0 &= R_c \mathbb{E}_{\mathcal{P}}[A] \\ \sigma_1 &= R_c \end{aligned} \tag{A.8}$$

By substituting the expressions above in (5) we obtain the variance bound in (5) in the form stated in the Proposition.

In order to show that the bound for $var(B)$ is equal to $var(B)$ we evaluate the latter explicitly using the Fluctuation-dissipation-theorem (see *e.g.* Lestas et al. (2008) and reference therein) noting that this is possible due to the linearity of the propensities.

■

A.3 Proof of Proposition 5

Proof.

By following similar arguments to the ones in A.1 and appropriate manipulation we can express the covariance of A, B as a Newton series of positive terms and obtain the statement.

■