Approximations of the CME

September 7, 2023

1 Description of the Birth-Death chemical master equation

The system consists of a molecular species with a count variable n. There are two reactions:

1. **Birth reaction:** Increases the molecule count by 1. Rate: $\lambda_b(n) = n \times \max[0, k + C_b(N_{ss} - n)]$
2. **Death reaction:** Decreases the molecule count by 1. Rate: $\lambda_d(n) = n \times k$

1.1 Gillespie Algorithm

The Gillespie algorithm can be used to simulate the stochastic dynamics of the system. The steps are:

1. Initialize the molecule count $n(0)$ at time $t = 0$.
2. At each time step, calculate the rates of the birth and death reactions:
   (a) $\lambda_b(n) = n \times \max[0, k + C_b(N_{ss} - n)]$
   (b) $\lambda_d(n) = n \times k$
3. Construct a discrete distribution for the two reactions, with probabilities:
   (a) $P[birth] = \lambda_b(n)/[\lambda_b(n) + \lambda_d(n)]$
   (b) $P[death] = \lambda_d(n)/[\lambda_b(n) + \lambda_d(n)]$

   And determine which reaction occurs by randomly drawing a sample. Update n accordingly.
4. Calculate the time until the next reaction by randomly drawing a sample from an exponential distribution with rate $\lambda = \lambda_b(n) + \lambda_d(n)$, and update time accordingly.
5. Return to step 2 until desired simulation time is reached.

2 Deriving the Langevin approximation to the CME

Consider a birth-death model where the birth rate k_b and death rate k_d are functions of n, i.e., $k_b(n)$ and $k_d(n)$. We can express the general model as an approximated Langevin equation:

$$dn = f(n)dt + g(n)\xi\sqrt{dt} \tag{1}$$

Where:

- $f(n)$ is the deterministic drift term representing the net change rate of n due to both birth and death reactions.
- $g(n)$ represents the strength of the noise term.
- $\xi\sqrt{dt}$ is a standard normal random variable driving the randomness of the system.

2.1 The drift term

2.1.1 Deriving $f(n)$

The deterministic drift term $f(n)$ is the expected change in n at each potential value of n. As such, it can be expressed as the negative difference between the expected number of birth and death events at that point.

$$f(n) = k_b(n) - k_d(n) = max[0, k + C_b \times (NSS - n)] \times n - k \times n \tag{2}$$

2.1.2 Exploring $f(n)$

The shape and magnitude of the drift term function $f(n)$ is highly dependant on two parameters of the equation system, namely the control strength C_b, and the basal rates of births and deaths k $(= k = k)$(Fig. 1).

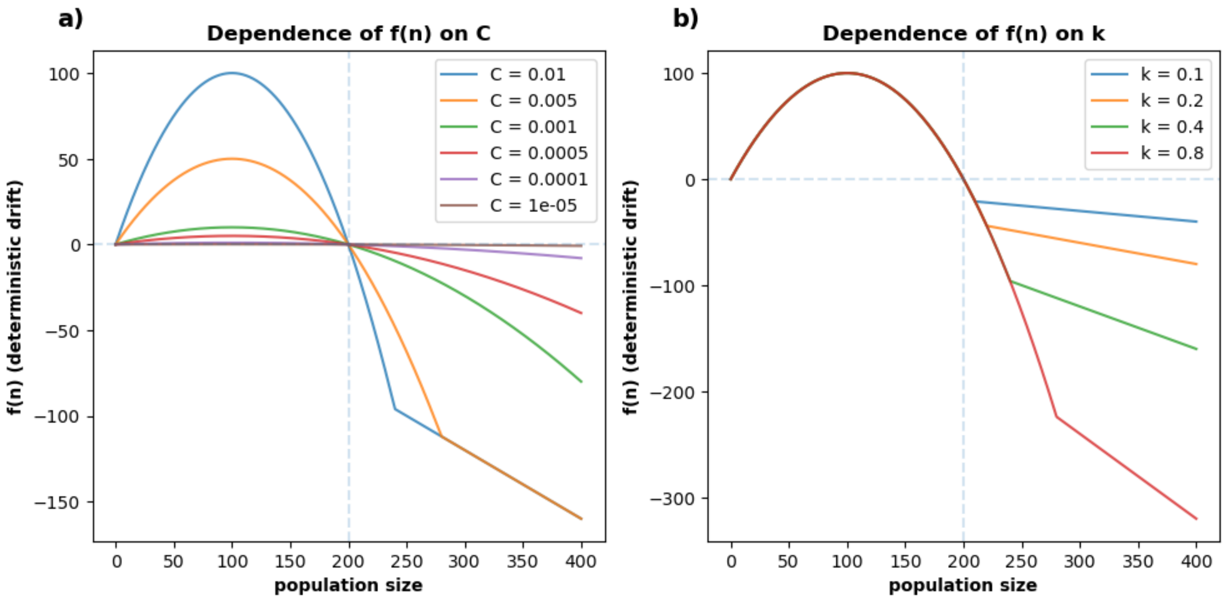

Figure 1: Dependence of drift term $f(n)$ on the control strength C and turnover k. **a)** $f(n)$ across various values of C_b when $k = 0.4$ **b)** $f(n)$ across various values of k when $C_b = 0.01$

2.2 The noise term $g(n)$

The $g(n)$ term in the Langevin equation captures the variance of the noise, i.e., the inherent stochasticity in the birth-death model.

2.2.1 deriving $g(n)$

Given that the fluctuations in n arise due to stochastic birth and death events, we can derive $g(n)$ by considering the variance in the changes of n during a small time interval dt.

Expectation and Variance of Δn

For a birth event, $\Delta n = +1$ occurs at rate $k(n)$ and for a death event, $\Delta n = -1$ occurs at rate $k(n)$. Therefore $E[\Delta n]$ is:

$$E[\Delta n] = (+1) \cdot k(n) \cdot dt + (-1) \cdot k(n) \cdot dt = (k(n) - k(n)) \cdot dt \tag{3}$$

Accordingly, the second moment is:

$$E[(\Delta n)^2] = (1^2) \cdot k(n) \cdot dt + (-1^2) \cdot k(n) \cdot dt = (k(n) + k(n)) \cdot dt \tag{4}$$

And the variance $\mathrm{Var}[\Delta n]$ in n during dt is:

$$\mathrm{Var}[\Delta n] = E[(\Delta n)^2] - [E[\Delta n]]^2 = (k(n) + k(n)) \cdot dt - [(k(n) - k(n)) \cdot dt]^2 \tag{5}$$

Substituting $k(n) = n \cdot max[0, k + C_b \cdot (NSS - n)]$ and $k(n) = n \cdot k$ yields:

$$\mathrm{Var}[\Delta n] = (n \cdot max[0, k + C_b \cdot (NSS - n)] + n \cdot k) \cdot dt - [n \cdot max[0, k + C_b \cdot (NSS - n)] - n \cdot k) \cdot dt]^2 \tag{6}$$

Which can be expanded and rearranged to give:

$$\begin{aligned}\text{Var}[\Delta n] =& n \cdot (\max[0, k + C_b(NSS - n)]dt + k)dt + \\ & n^2 \cdot (2k \max[0, k + C_b(NSS - n)] - k^2 - \max[0, k + C_b(NSS - n)]^2)dt^2\end{aligned} \tag{7}$$

A suitable linear approximation of Var[Δn] is therefore:

$$\text{Var}[\Delta n] \cong n \cdot (\max[0, k + C_b(NSS - n)]dt + k)dt \tag{8}$$

Deriving $g(n)$ from $Var[\Delta n]$

In the Langevin equation, the term $g(n)\xi\sqrt{dt}$ represents the stochastic or random change in n over that same time increment. Thus the variance of Δn can be expressed as:

$$\Delta n = g(n)\xi\sqrt{dt} \implies Var[\Delta n] = Var[g(n)\xi\sqrt{dt}] \tag{9}$$

Since ξ is a standard normal random variable with variance 1:

$$\text{Var}[\Delta n] = \text{Var}[g(n)\xi\sqrt{dt}] = g(n)^2 \text{Var}[\xi]dt = g(n)^2 dt \tag{10}$$

Substituting back eq. 8 and diving by dt:

$$g(n)^2 = \text{Var}[\Delta n]/dt = \frac{n \cdot (\max[0, k + C_b(NSS - n)]dt + k)dt}{dt} = n \cdot (\max[0, k + C_b(NSS - n)]dt + k) \tag{11}$$

Taking the square root finally gives:

$$g(n) = \sqrt{n \cdot (\max[0, k + C_b(NSS - n)]dt + k)} \tag{12}$$

2.2.2 Exploring $g(n)$

The shape and magnitude of the noise term $g(n)$ is dependant on two parameters of the equation system, namely the control strength C_b, and the basal rates of births and deaths k (= k = k)(Fig. 2).

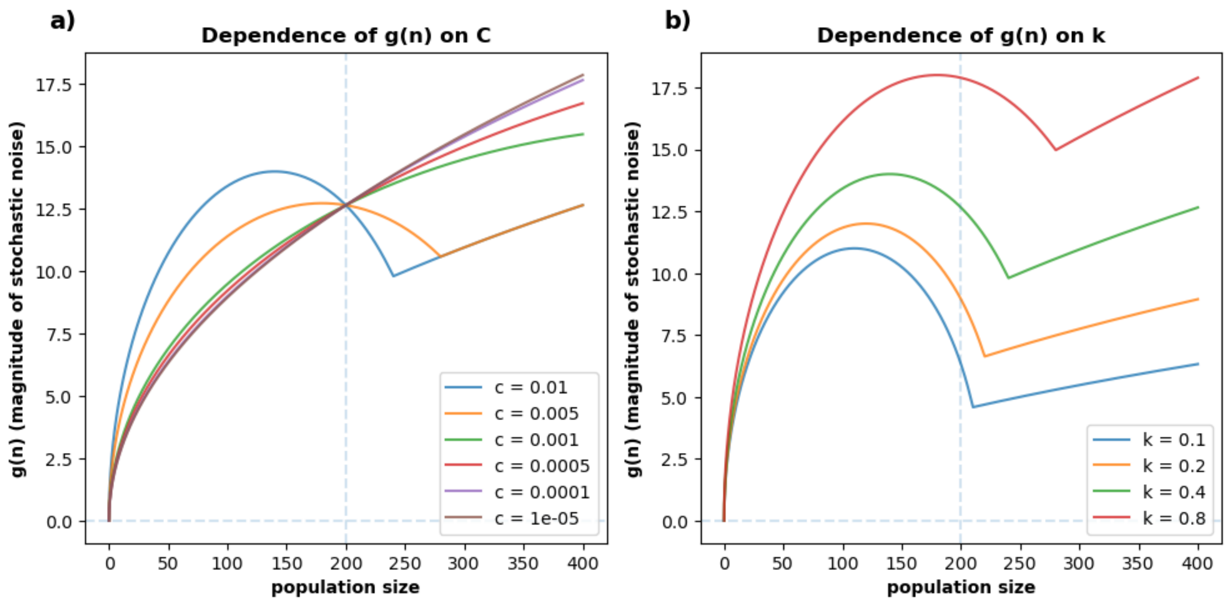

Figure 2: Dependence of the noise term $g(n)$ on the control strength C and turnover k. **a)** $g(n)$ across various values of C_b when $k = 0.4$ **b)** $g(n)$ across various values of k when $C_b = 0.01$

.

2.3 Complete equation

Combinding the deterministic drift term, and the stochastic change yields the complete SDE describing the change in population size over time:

$$dn = n \cdot (max[0, k + C_b \cdot (NSS - n)] - k)dt + \sqrt{n \cdot (\max[0, k + C_b(NSS - n)]dt + k)} \cdot \xi\sqrt{dt} \tag{13}$$

2.4 Simulation using the Euler-Maruyama method

We wish to approximate the SDE defined by eq. 13. Using the Euler-Maruyama method.

$$\Delta n = f(n)\Delta t + g(n)\xi\sqrt{\Delta t} \tag{14}$$

2.4.1 Update equation

When the system is simulated using the Euler-Maruyama method, the drift term becomes:

$$f(n)\Delta t = n \cdot (max[0, k + C_b \cdot (NSS - n)] - k)\Delta t \tag{15}$$

The magnitude of the noise term is calculated as:

$$g(n) = \sqrt{n \cdot (max[0, k + C_b \cdot (NSS - n)] + k)} \tag{16}$$

Which is then multiplied by a sample taken from a standard normal distribution, and the square root of the timestep to yield the noise term:

$$g(n)\xi\sqrt{\Delta t} = \sqrt{n \cdot (max[0, k + C_b \cdot (NSS - n)] + k)} \cdot \xi\sqrt{\Delta t} \tag{17}$$

Therefore the complete update equation for a discrete time step is:

$$n_{t+\Delta t} = n_t + n(max[0, k + C_b \cdot (NSS - n)] - k)\Delta t + \sqrt{n(max[0, k + C_b \cdot (NSS - n)] + k)} \cdot \xi\sqrt{\Delta t} \tag{18}$$

3 Continuous-Time Markov Chain representation of the CME

In stochastic modelling, a birth-death process can be approximated by a Continuous-Time Markov Chain (CTMC), due to the memoryless property, the natural incorporation of time-dependent transition rates, and the discretized state space of the original birth-death process.

In a birth-death process, entities can "birth" into or "die" out of various states at specific rates. These rates can be encapsulated in a transition rate matrix Q, which provides a compact and mathematically tractable way to describe the system's dynamics. The CTMC framework allows for the analytical solution of various performance metrics, such as state probabilities through the use of differential equations and matrix exponentials.

Moreover, CTMCs offer a high degree of flexibility in modeling different kinds of birth-death processes, including those with state-dependent rates, multiple types of transitions, and more complex state spaces. Therefore, the CTMC serves as a robust and versatile framework for approximating birth-death processes, providing both conceptual clarity and analytical solutions.

3.1 Defining the state space S

Due to the way in which birth rates are specified, there are two boundaries in the state space of the birth-death process. The first trivial boundary is the absorbing state $n_{min} = 0$. The second boundary occurs when the birth rate drops to 0, above which the population size cannot grow.

$$k_b(n) = 0 \implies n \cdot max[0, k + C_b \cdot (NSS - n)] = 0 \implies n_{max} = NSS + \frac{k}{C_b} \tag{19}$$

Therefore, the corresponding set of all states for the CTMC is $S = \{0, 1, 2, ..., n_{max}\}$, and the transition rate matrix Q will have size $n_{max} + 1$

3.2 Defining the transition rate matrix Q

As Each state i of the CTMC corresponds to a given population size n in the birth-death process, transitions between different states of the Markov Chain $q_{i,j}$ correspond to singular birth and death events which change the population size by 1. As discussed previously, these the rate of these events (λ_b, λ_d) themselves are dependant on the population size at which the events occur $(\lambda_b(n), \lambda_d(n))$. Therefore, the transition rate matrix Q is defined as:

$$Q = \left(q_{ij}\right)_{0\leq i,j\leq N} \text{ where } \begin{cases} q_{0,j} = 0, & \text{for all } j \text{, since } i = 0 \text{ is absorbing} \\ q_{i,j=i+1} = \lambda_b(i) = i \cdot max[0, k + C_b \cdot (NSS - i)], & \text{for } i > 0 \text{, the rate of birth events} \\ q_{i,j=i-1} = \lambda_d(i) = i \cdot k, & \text{for } i > 0 \text{, the rate of death events} \\ q_{i,j} = 0, & \text{otherwise} \end{cases} \tag{20}$$

To satisfy the properties of transition rate matrices, the additional constraint $q_{ii} = -\sum_{j\neq i} q_{ij}$ is also applied to give the tridiagonal matrix:

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ \lambda_d(1) & -q_1 & \lambda_b(1) & 0 & \cdots & 0 \\ 0 & \lambda_d(2) & -q_2 & \lambda_b(2) & \cdots & 0 \\ 0 & 0 & \lambda_d(3) & -q_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_d(n_{max}) & -\lambda_d(n_{max}) \end{pmatrix} \tag{21}$$

3.3 Defining the transition probability matrix over time $P(t)$

A fundamental result in the theory of CTMCs arises from the solution to the continuous-time Chapman-Kolmogorov equation. Given the infinitesimal generator matrix Q, the transition probability matrix at time t can be obtained via a matrix exponential:

$$P(t) = e^{Q\cdot t} \tag{22}$$

3.4 Calculating the distribution of states over time $\pi(t)$

The discrete distribution $\pi(t)$ corresponds to the state space S of the CTMC. Each element $\pi_i(t)$ represents the probability of the system being in state i at time t. $\pi(t)$ is determined by the initial state distribution $\pi(0)$ and the transition probability matrix $P(t)$ according to:

$$\pi(t) = \pi(0) \times P(t) \implies \pi(t) = \pi(0) \times e^{Q\cdot t} \tag{23}$$

4 Results

Simulations comparing the Gillespie system and the Langevin model show extremely close correspondence. At higher values of C, the approximation successfully recreates the slight but stable drop in population size (Fig. 3), while at lower C values, the approximation successfully follows the exact decreasing trajectory (Fig. 4). As C nears 0, the approximation again recreates the lack of change in n over time (Fig. 5).

5 Interpretation

These results suggest that the peculiar non-monotonic behaviour of $f(n,t)$ with respect to changes in C_b observed in the Gillespie SSA are the result of the interaction between the restoring force $f(n)$, and the amount of stochastic variation $g(n)$.

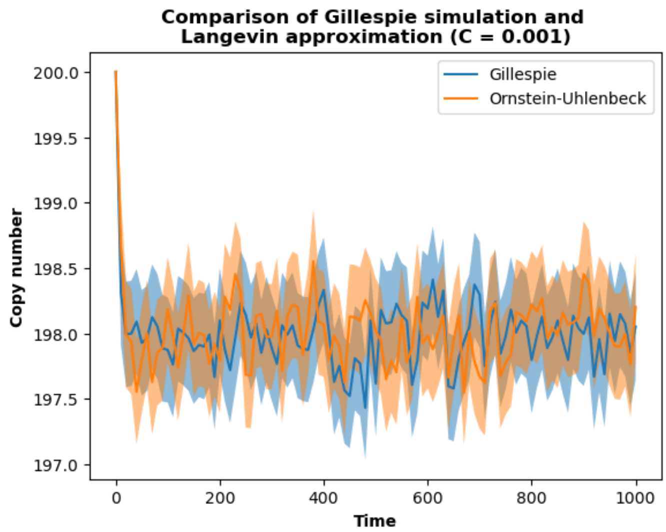

Figure 3: Comparison of copy numbers over time at a high C value. (10000 replicates, $k = 0.4$)

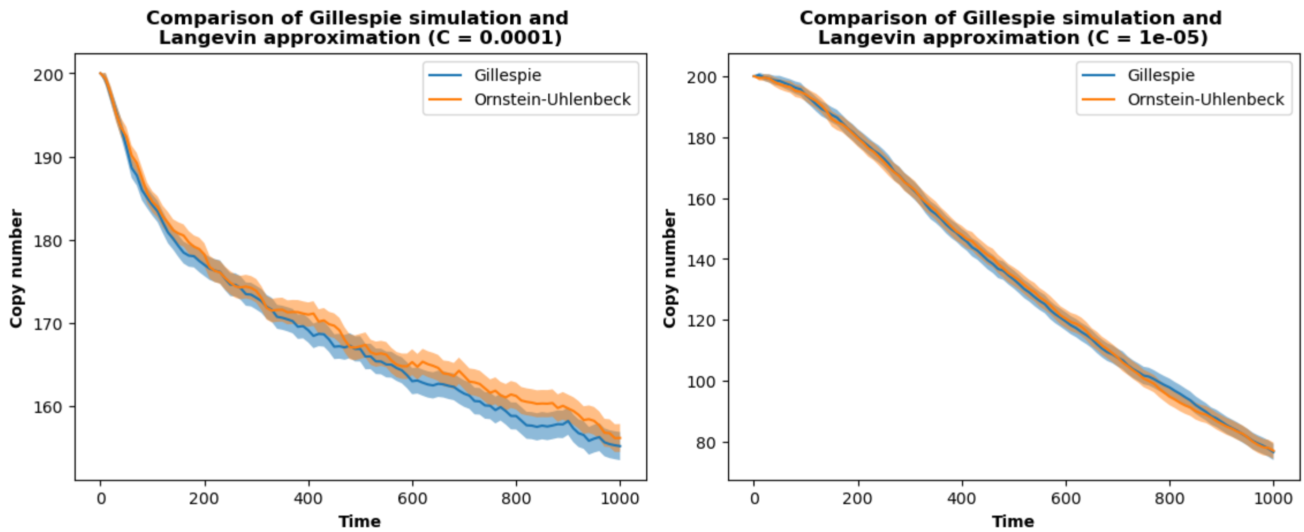

Figure 4: Copy numbers decrease over time at lower C values (10000 replicates, $k = 0.4$)

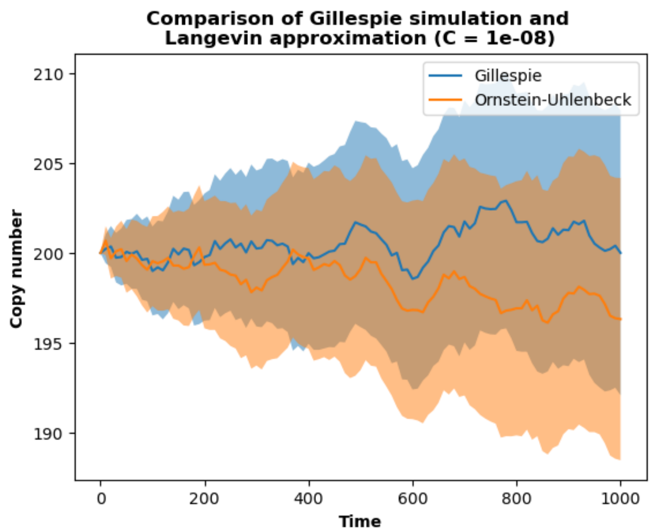

Figure 5: Copy numbers are stagnant over time when $C \approxeq 0$ (10000 replicates, $k = 0.4$)