

## Notes

### Abstract

- Natural systems are heterogenous.
- Mathematical models usually ignore heterogeneity.
- stochastic methods account for heterogeneity
  - based on the fact that systems inherently contain at least one source of heterogeneity (intrinsic heterogeneity)
- three different sources of heterogeneity
- discrete-state stochastic methods track only populations
  - assume the volume is spatially homogeneous

### 1. Introduction

systems biology

- looking at systems-level dynamics instead of one component at a time

Model

- an abstract representation of the system, usually formulated mathematically.
- however, analytical solutions to simple models do exist and we need to solve the model *numerically*

Numerical solutions are simulations

- trying to mimic the behavior of the real system over time

All models are not reality but abstractions of reality, simplified versions of the systems they represent.

Almost all models are phenomenological.

Different modelling/simulation methods:

- most accurate and most computationally expensive - molecular dynamics
  - Quantum methods - evaluating wavefunctions
  - Classical methods - solve classical equations of motion for molecules to solve their motion deterministically

Several classes of stochastic methods:

- continuous deterministic methods (ODE and PDE)
- several transitions bet. random and deterministic methods

Classical molecular dynamics can simulate huge numbers of discrete particles for very short simulation times.

ODE look only at concentrations of macromolecules but simulate for longer time.

Paper: Overview of discrete-state stochastic simulations (the time variable is continuous)

- simulate molecular populations over relatively long time periods, while still regarding them as discrete units
- they lose their spatial aspect and must assume that the system is stochastic

### 2. Heterogeneity

three main sources:

1. genetic (nature)
  - not the only means of difference in expression
  - counterexample: convergent evolution
    - different species, same phenotype
  - convergence of phenotype with a corresponding convergence in genotype
2. environmental (nurture, extrinsic noise)
  - broad term indicating neither genetic or intrinsic
3. stochastic (change, intrinsic noise)
  - arises from random thermal fluctuations at the level of individual molecules
  - only random fluctuations are inherent to the expression of a single gene
  - noise affects the expression of proteins differently
  - key in cellular decision making
  - plays a crucial role in causing genetic mutations

Cellular decision making involves both environmental and intrinsic heterogeneity

Evolution is a powerful example of the interplay of all three.

### 3. Deterministic vs. stochastic models

Continuous Deterministic simulation

- differential equations
- continuous
- do not include noise
- accurate when interested in the mean dynamics
- works when system is large (approx. molecules as concentrations)
- can only find one point in the distribution
- generally faster than stochastic

Discrete Stochastic models

- simulates actual molecular numbers
- used for small populations where intrinsic noise increases
- many simulations reveals the full probability distribution

Deterministic models are complementary to the stochastic methods

an intermediate between the two are continuous stochastic models

- relatively large systems to consider concentrations but still include noise

#### 4. Accounting for stochasticity

- focus only on how the total number of each molecular species changes through time and label all extra information as blackbox represented as randomness
- treat the problem as stochastic
- two key assumptions for non-spatial stochastic methods
  1. non-reactive collisions are much more frequent than reactive ones
  2. the system is in thermal equilibrium
- both ensure a well-mixed system
- the time between reactions has an exponential distribution
- the number of reactions occurring in some time interval is given by a Poisson random number.
- allows the use of mass action kinetics
  - propensities are given by a rate constant multiplied by the abundance of reactants
- with these conditions, a Markov jump model can be used to model the time evolution
- results in huge decrease in computational effort while being statistically exact
- this exactness may not be a huge factor due to
  1. the model not being exact representation of reality
  2. the model is statistically exact at the limit of infinite simulations
  3. uncertainty in measurements result to errors
- assumes that reactions are activation controlled
  - molecules diffuse to within reacting distance of each other faster than the reaction can occur
- non-spatial stochastic methods remain useful even more for systems with low number of molecules.

#### 5. Stochastic Methods

- if the conditions are satisfied
  - well mixed
  - thermal equilibrium
  - molecules moving randomly
  - probability of reaction depends on the number of reactants
- then we can model the time evolution as a Markov jump process

- genetic heterogeneity can be set at the start and kept constant (constant throughout simulation time)
- environmental heterogeneity can be incorporated by varying propensities

#### Master equation

- an exact and full description of the behavior of the Markov process through time.
- coupled DE (one for each possible state of the system)
- two contributions
  - probability that already in state  $X$  and did not change
  - probability that in another state and changed into  $X$
- derives the moment equations for stochastic model
- disadvantage
  - complexity
  - analytical solution only for very simple systems with linear reaction propensities
- application of linear noise approx. of van Kampen
  - yields information on the behaviour of the mean plus small fluctuations about it
- alternative are **trajectory-based approaches**
  - simulate single realization

#### Stochastic simulation algorithm

- statistically exact
- generates full probability distribution identical to Markov process
- steps along time:
  - when the next reaction will happen
  - which reaction will happen next
- limitation:
  - slow, must simulate every reaction
  - becomes slow for many reactions or larger populations

#### Tau-Leaping

- single stochastic realisation from the full distribution
- much larger steps than SSA
- counts total number of occurrences of each type reaction
- faster than SSA but problem dependent
- Drawback: lost accuracy compared to SSA
  - we do not know exactly how many reactions occurred
  - we do not know exactly when each reaction occurred between the time step

- reactions may affect the probability of other reactions
- may be computationally intensive,  $M$  poisson numbers
- the step size can be chosen to maximize computational efficiency and accuracy
- another issue: negative reactions due to molecular species being depleted between time steps

### Higher-order tau-leap

order of accuracy - decrease in error for every decrease in time step

- usually 1 in tau-leaping
- but methods exist that have order greater than one
- disadvantage is mathematically more complicated

### Multiscale methods

for cases where reactions operate at different time scales (slow and fast reactions) we can use:

1. implicit methods
  - based on deterministic methods for stiff systems
  - use normalized time steps by expanding the range of time steps where the method is stable
2. multiscale algorithms
  - partitions reactions into fast and slow types and simulates both separately
  - fast reactions are simulated using an approximate method
    - assuming that slow reactions are constant
  - slow reactions use stochastic methods
    - assume fast reactions have relaxed
  - results to significant reductions in computational time
  - but introduces errors from the coupling of the two scales and errors in simulating the fast reactions

### Summary and Outlook

focus limited to discrete stochastic methods and do not take spatial considerations into account

Paper opinion: three areas where we could make significant improvements:

1. parallel stochastic methods (parallel computing and GPU computing)
2. multiscale methods
3. spatial stochastic methods

### Main Idea(s) of the Paper

In sentence form

1. There are three main sources of heterogeneity in biological systems.
  - Genetic heterogeneity is based on the nature of the system but does not account for all heterogeneity in a biological system.
  - Environmental sources is a broad terms that indicates neither of the other two sources, but is also accounted for in most biological models.
  - Stochastic or intrinsic heterogeneity account for many differences in biological systems but is rarely included in biological systems models.
- 2.

### Implications: the dynamics of complex living systems

1. implications of this paper's ideas to our understanding of the dynamics of complex living systems?

What are the strong points?

What are the weak points?

2. most unique or new feature of the technique or method used?

at least one difference from another approach to the same problem

3. features that may be useful in other approaches. How do you think can this be done?

### Personal impression

How is the paper useful for you? Why, or why not?

Is there anything new to you in it?

Do you find anything that can be improved in the presentation/graphs/illustrations?

There is a long definition of terms part in the introduction that I find unnecessary considering the main point of the paper is only on discrete-state stochastic simulations.

Most diagrams are conceptual images and do not portray experimental results. As they were already discussed in the body, I find them unnecessary.

Is there any statement or claim of the author(s) that you don't agree with? Explain your position.

### Appendix