Integrating the Gillespie Algorithm with Swarm

Scott Christley
Dept. of Mathematics, Dept. of Computer Science
Center for Mathematical and Computational Biology
Center for Complex Biological Systems
University of California Irvine

Motivation (Systems Biology)

- ABM biological cell behaviors
  - Growth, movement, division, death
  - Interactions: cell-cell and cell-environment
  - Individual cell behaviors produce emergence of population level behavior (tissue function).
- Add more detail, selectively
  - Cell-cell interaction (Notch signaling pathway)
  - Insert gene network within each cell to “drive” one of the cell behaviors
  - Leave the other cell behaviors as ABM actions

Motivation (Systems Biology)

- Matching time scales (spatial scales)
  - Multi-scale, behaviors operating at different time scales
- Gene network simulation
  - Ordinary Differential Equations
    - Continuous, deterministic, population averages
    - Easier: numerically integrate over any time frame
    - Implicit separation of time scales (morpho-static limit)
  - Stochastic chemical kinetics
    - Exact for low molecule counts
    - Strong statistical correlations can be formed

Stochastic Chemical Kinetics

- Rule-based description

$$\varnothing \xrightarrow{k_1} X$$ Zero order

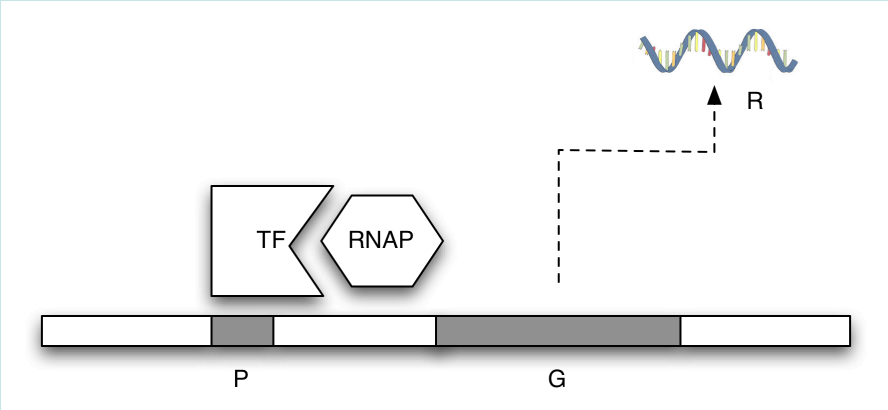
$$X \xrightarrow{k_1} \varnothing$$ First order

$$X + Y \xrightarrow{k_1} Z$$ Second order

$$X + X \underset{k_2}{\overset{k_1}{\Leftrightarrow}} X_2$$

Stochastic Chemical Kinetics

- Gene transcription



$$TF + P \Leftrightarrow TF \circ P$$

$$TF \circ P + RNAP \Leftrightarrow TF \circ P \circ RNAP$$

$$TF \circ P \circ RNAP + G \rightarrow TF \circ P + RNAP + G + R$$

Gillespie Algorithm (GSA)

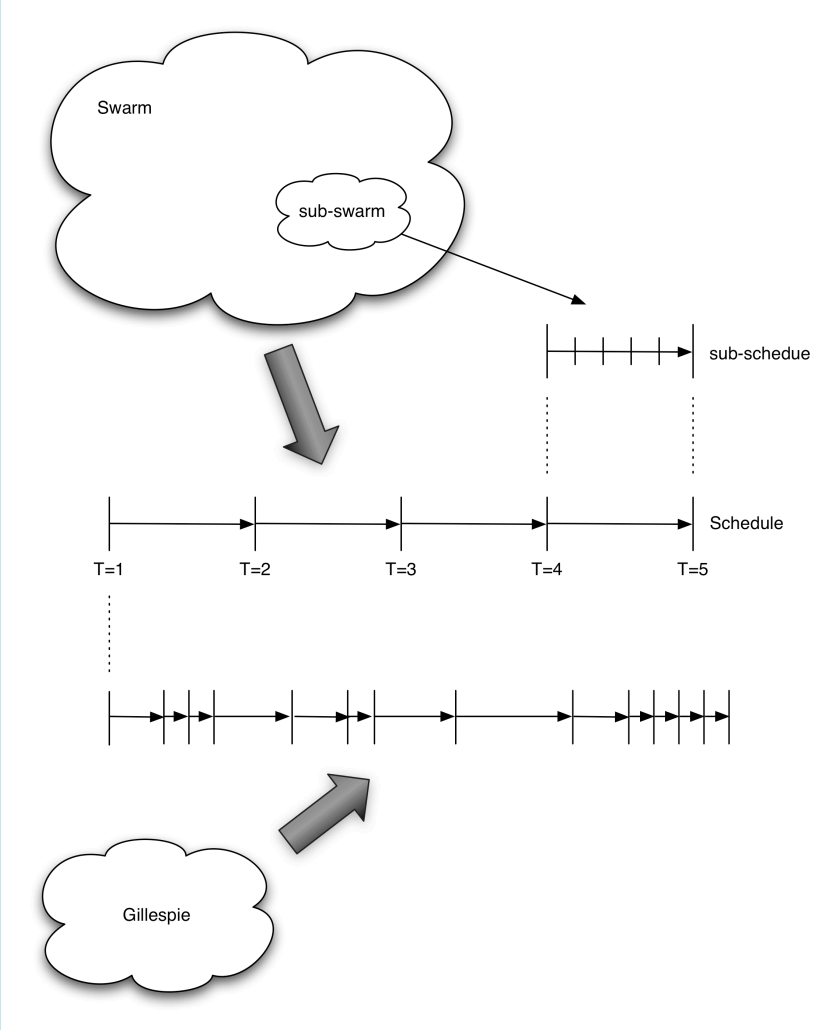
- Exact simulation of stochastic chemical kinetics (Gillespie 1977)
- Calculate propensity functions for each rule, which are normalized into probabilities
  - The larger the molecule count, the greater propensity for the rule.
- Rules occur at a rate per unit time according to Poisson distribution (basis in physical law).
  - However, non-homogenous Poisson process because rate changes with quantity of molecules
- Time of next reaction: Exponential distribution
- Randomly pick one reaction based upon rule probabilities
- Execute rule: update molecule counts
- Recalculate rule probabilities
- Repeat

Gillespie Algorithm (GSA)

- Similar to agent-based simulation
  - Discrete entities (molecules)
  - Probabilistic rules of behavior
- Key differences
  - Entities (molecules) are not distinct agents, it is assumed that they indistinguishable from each other.
  - Thus only need to maintain counts (totals).
  - Time move forwards in random steps (exponential distribution)
  - Time scale is defined by the number of molecules and rule rates
    - Many molecules/fast rates: small time scale
    - Few molecules/slow rates: large time scale

Matching Time Scales

- ABM: discrete time
  - Consistent sub-intervals for sub-swarms
- GSA: random continuous time
- If this was the only issue, then integration can be handled:
  - Run GSA
  - If time for next reaction exceeds ABM next time
  - Then save next reaction time, perform ABM rules, go back to GSA



Other Problems

- ABM rules based upon molecule count
  - If (X > 100) then perform ABM rule
  - Molecule counts can fluctuate greatly within a unit time interval, events can be missed if only check at time boundary
- How long does an ABM rule "take"?
  - ABM typically assume each rule is performed instantaneously.
  - Biological processes like cell division, movement, etc. are complicated tasks that can take a long time compared individual chemical reactions.

Solution 1

- Approximate GSA
  - Convert GSA to unit time (time discretization)
  - Others: Tau-leap method, Langevin equation
- Time Discretization
  - Do not simulate each individual reaction.
  - Random draw from Poisson distribution for each reaction
  - Count of occurrence for each reaction
  - Apply all rules at each unit time

Time Discretization (Solution 1)

- Bring GSA "up" to ABM time scale

- Pros
  - Computationally efficient
- Cons
  - Potentially lose stochastic nature
  - Inaccuracy increases with longer time intervals
  - Invalid conditions: negative molecule counts

Solution 2

- Bring ABM “down” to GSA time scale

- Rewrite ABM rules just like GSA rules with appropriate rates.
- Just run Gillespie algorithm for everything
- No more discrete time, simulation is completely random continuous time

Solution 2

- Bring ABM "down" to GSA time scale
- Pros
  - It is "exact" (though ABM might not be Poisson)
- Cons
  - Computationally expensive
  - Can be hard to determine ABM rule rates
  - ABM rules can essentially become rare events
  - Somewhat defeats the purpose of detailed interactions "driving" an ABM rule

Solution 3

- Both solution 1 and 2 attempt to collapse the multi-scale problem into single time scale.

- Time Integral GSA
  - Normal GSA, simulate each reaction
  - Estimate the likelihood of a random variable based upon the reactions performed during time step.

$$L(r;X,C) = \frac{1}{Z}\int_{t_0}^{t_1} p_k(X,C)dT = \frac{1}{Z}\sum_{i=0}^{n_i} p_k(X,C)\left[t_i - t_{i-1}\right]$$

- Rewrite ABM rule
  - If (random[0,1] < r) then perform ABM rule

Solution 3

- Time Integral GSA
- Cons
  - Might not be obvious exactly what integral to calculate, or proper normalization.
- Pros
  - Many ABM rules have probabilistic rate anyways, so this ties that probability to a detailed underlying stochastic process.
  - Agent heterogeneity introduced by stochastic process that can change over time, versus random initial condition.

Summary

- Multi-scale requires careful consideration.
  - Interaction between scales
  - Just considered temporal, spatial introduces additional issues
- Techniques that attempt to collapse scales suffers from problems.
  - “up” is approximate, lose stochastic detail
  - “down” is computationally expensive
- Maintain the separation of scales. Define a functional relationship between the lower level process and the parameter (decision process) at the higher level.
  - Time Integral GSA
- Issue of non-instantaneous higher level behaviors is not resolved, not an issue if behavior is independent of lower level.
  - Cell movement, non-movement related intercellular process