Problems with Reaction Rate Equations

- Chemical reaction network model can be transformed using law of mass action into ODEs known as reaction rate equations.
- Assume concentrations vary continuously and deterministically.
- Chemical systems satisfy neither of these assumptions.
- Number of molecules of a species is a discrete quantity.
- Chemical reactions occur after two molecules collide.
- Unless track exact position and velocity of every molecule, not possible to know when a reaction may occur.
- Should consider occurrence of reactions to be stochastic.

Stochastic Process Description

- For systems which involve large molecular counts, ODE models give accurate picture of their behavior.
- If molecular counts are small, discrete and stochastic nature may have significant influence on observed behavior.
- Genetic circuits typically involve small molecule counts.
- Often only one strand of DNA and a few 10s or 100s of molecules of each transcription factor.
- Accurate analysis requires a stochastic process description.
- This lecture presents one such description, the chemical master equation, and algorithms to analyze it.

A Stochastic Chemical Kinetic Model

- Composed of $n$ chemical species $\{S_1, \ldots, S_n\}$ and $m$ chemical reaction channels $\{R_1, \ldots, R_m\}$.
- Assume species contained within constant volume $\Omega$.
- Assume system is well-stirred to neglect spatial effects.
- Assume system is in thermal equilibrium (i.e., at a constant temperature), but not necessarily chemical equilibrium.

State Updates

- $X_i(t)$ is the number of molecules of $S_i$ at time $t$.
- $X(t) = (X_1(t), \ldots, X_n(t))$ is the state of a system at time $t$.
- $X(0) = x_0$ is initial number of molecules at initial time $t_0$.
- After $R_p$, the new state is $x' = x + v_p$, where $v_p = (v_{p1}, \ldots, v_{pn})$ is the state-change vector and $v_{ji}$ is the change in $S_j$ due to $R_p$.
- The 2-dimensional array $\{v_{ji}\}$ is known as the stoichiometric matrix.
- $R_p$ is elemental if it can be considered a distinct physical event that happens nearly instantaneously.
- For elemental $R_p$, values of $v_{ji}$ are constrained to $0, \pm 1, \pm 2$.

Specific Probability Rate Constant

- Every $R_p$ has a specific probability rate constant, $c_p$, which is related to the reaction rate constant, $k_p$.
- $c_p dt$ is the probability that a randomly chosen combination of reactant molecules react as defined by $R_p$ inside $\Omega$ in $[t, t + dt]$.
- Multiplying $c_p$ by the number of possible combinations of reactant molecules for $R_p$ in a state $x$ yields the propensity function, $a_p$.
- $a_p(x) dt$ is the probability that $R_p$ occurs in the state $x$ within $\Omega$ in the next infinitesimal time interval $[t, t + dt]$.
A Bimolecular Reaction Channel

- A typical bimolecular reaction channel $R_\mu$ has form:
  \[ S_1 + S_2 \xrightarrow{c_\mu} S_3 + \ldots \]
- $c_\mu$ is probability that a $S_1$ molecule and $S_2$ molecule collide and react within next $dt$ time units.
- Assume molecules hard spheres with masses $m_i$ and radii $r_i$.
- Thermal equilibrium means that a selected $S_i$ can be found uniformly distributed within $\Omega$.
- Also means avg. relative speed in which $S_2$ sees $S_1$ moving is:
  \[ \mathcal{V}_{12} = \sqrt{8k_B T / \pi m_{12}} \]
  where $k_B$ is Boltzmann’s constant and $m_{12} = m_1 m_2 / (m_1 + m_2)$.

A Bimolecular Reaction Channel (cont)

- Number of combinations of $S_1$ and $S_2$ molecules is $x_1 x_2$, so propensity function for $R_\mu$ is $a_\mu(x) = c_\mu x_1 x_2$.
- If $S_1 = S_2$ then number of combinations is $x_1 (x_1 - 1) / 2$, and $a_\mu(x) = c_\mu x_1 (x_1 - 1) / 2$.

Monomolecular Reactions

- Monomolecular reactions are of this form:
  \[ S_1 \xrightarrow{c_\mu} S_2 + \ldots \]
- $S_1$ makes a spontaneous change in its internal structure.
- $c_\mu$ must be found from quantum mechanical considerations.
- Propensity function is simply $a_\mu(x) = c_\mu x_1$.
- If it is actually an enzymatic reaction of the form:
  \[ E + S_1 \xrightarrow{c_\mu} E + S_2 + \ldots \]
  where $E$ is an enzyme, should be considered as a bimolecular reaction.

Trimolecular Reactions

- Trimolecular reactions are of this form:
  \[ S_1 + S_2 + S_3 \xrightarrow{c_\mu} S_4 + \ldots \]
- Probability is very small, so typically used as approximation for:
  \[ S_1 + S_2 \xrightarrow{c_\mu} S' \] and $S' + S_3 \xrightarrow{c_\mu} S_4 + \ldots$
- This approximation is reasonable when the lifetime of $S'$ is very short (i.e., $1/c_\mu$ is very small).
- The probability that a molecule of $S'$ reacts with a randomly chosen molecule of $S_3$ is approximately $c_\mu (1/c_\mu)$.

Trimolecular Reactions (cont)

- Consider a small but finite time interval $\Delta t$ which is still much larger than the lifetime of $S'$ (i.e., $\Delta t >> 1/c_\mu$).
- If $\Delta t$ is sufficiently small, then the probability that $S_1$ and $S_2$ react in that time interval to form $S'$ is $c_\mu \Delta t$.
- Probability of both reactions occurring in $\Delta t$ is $(c_\mu c_\mu / c_\mu) \Delta t$, so $c_\mu$ for the trimolecular reaction approximation is:
  \[ c_\mu' = c_\mu c_\mu / c_\mu \]
- Approximation because $\Delta t$ is not a true infinitesimal.
- Propensity function for this reaction is $a_\mu(x) = c_\mu x_1 x_2 x_3$. 
Relationship Between $c_\mu$ and $k_\mu$

- For bimolecular reactions, $c_\mu$ is proportional to $\Omega^{-1}$.
- For monomolecular reactions, it is independent of volume.
- For trimolecular reactions, it is proportional to $\Omega^{-3}$.
- In general, if $m$ is the number of reactant molecules in $R_\mu$:
  $$c_\mu \propto \Omega^{-(m-1)}$$
- Key to understanding relationship between $c_\mu$ and $k_\mu$.
- For monomolecular reactions, $c_\mu$ is equal to $k_\mu$.
- For bimolecular reactions, $c_\mu$ is equal to $k_\mu/\Omega$ if the reactants are different species and $2k_\mu/\Omega$ if the same species.

Time Evolution of Probability

- Not possible to know the exact state $X(t)$.
- Only can know probability of being in a given state at time $t$ starting from a state $X(t_0) = x_0$ (i.e., $P(x,t|x_0,t_0)$).
- Probability using a time-evolution of step $dt$ is shown below:
  $$P(x,t+dt|x_0,t_0) = P(x,t|x_0,t_0) \times \left[ 1 - \sum_{j=1}^{m} (a_j(x)dt) \right] + \sum_{j=1}^{m} P(x-v_j,t|x_0,t_0) \times (a_j(x-v_j)dt).$$
  - $dt$ is small enough that at most one reaction occurs during $dt$.

Chemical Master Equation

- Chemical master equation (CME) defines time evolution of state probabilities, $P(x,t|x_0,t_0)$:
  $$\frac{dP(x,t|x_0,t_0)}{dt} = \lim_{dt \to 0} \frac{P(x,t+dt|x_0,t_0) - P(x,t|x_0,t_0)}{dt} = \sum_{j=1}^{m} a_j(x-v_j)P(x-v_j,t|x_0,t_0) - a_j(x)P(x,t|x_0,t_0).$$
  - Typically cannot be solved analytically since it represents a set of equations as large as the number of molecules in the system.

Stochastic Simulation

- Trajectories for $X(t)$ can be generated using stochastic simulation.
- Could pick a small time step $dt$ and at each step update the system state by selecting a reaction to occur or doing nothing.
- For a sufficiently small $dt$, however, the vast majority of time steps result in no reaction.

Gillespie's Stochastic Simulation Algorithm

- Gillespie's stochastic simulation algorithm (SSA) improves the efficiency of simulation by stepping over useless time steps.
- Not based directly on CME, but equivalent form that uses $\rho(\tau, \mu, x, t)$.
- Defined such that $\rho(\tau, \mu, x, t)dt$ is probability that the next reaction is $R_\mu$ which occurs in $[t+\tau, t+\tau+dt]$ assuming current state is $X(t) = x$.
- This is a joint PDF for two random variables, $\tau$ and $\mu$ given that the system is in state $x$ at time $t$.
- Simulation advances from one reaction to the next skipping over time points in which no reaction occurs.
Derivation of Gillespie’s SSA

- Introduce \( P_0(\tau |x, t) \) that represents probability that there is no reaction in the time interval \([t, t + \tau] \).
- \( p(\tau, \mu|x, t) \) defined as follows:
  \[
p(\tau, \mu|x, t) = P_0(\tau|x, t) \times (a_\mu(x) \, dt). \tag{1}
\]
- No reactions occur in the interval \([t, t + \tau] \) and the \( R_\mu \) reaction occurs in the interval \([t + \tau, t + \tau + dt] \).

Derivation of Gillespie’s SSA (cont)

- Inserting Equation 2 into Equation 1 and canceling \( dt \) yields:
  \[
p(\tau, \mu|x, t) = \exp(-a_\mu(x)\tau) \times a_\mu(x),
\]
  which can be rewritten as:
  \[
p(\tau, \mu|x, t) = a_\mu(x) \exp(-a_\mu(x)\tau) \times \frac{a_\mu(x)}{a_\mu(x)}.
\]
- \( p(\tau, \mu|x, t) \) can be divided into PDFs for \( \tau \) and \( \mu \).
- \( \tau \) is exponential random variable with mean and std dev of \( \frac{1}{a_\mu(x)} \).
- \( \mu \) is integer random variable with point probabilities \( \frac{a_\mu(x)}{\sum_j a_j(x)} \).

Gillespie’s SSA (Direct Method)

- Initialize: \( t_0 \) and \( x = x_0 \).
- Evaluate \( a_\mu(x) \) and \( a_0(x) = \sum_{j=1}^{\infty} a_j(x) \).
- Draw two uniform random numbers, \( r_1 \) and \( r_2 \).
- Determine the time, \( \tau \), until the next reaction:
  \[
  \tau = \frac{1}{a_0(x)} \ln \left( \frac{1}{r_1} \right).
\]
- Determine the next reaction, \( \mu \):
  \[
  \mu = \text{the smallest integer satisfying } \sum_{j=1}^{\mu} a_j(x) > r_2 a_0(x).
\]
- Determine the new state: \( t = t + \tau \) and \( x = x + \nu_\mu \).
- If \( t \) is greater than the desired simulation time then halt.
- Record \((x, t)\) and goto step 2.

Simulation of \( PRE \) and \( OR \) Promoters

![Comparison of SSA to ODE (OR=1, PRE=1, RNAP=30, P1=35)](image)

Discussion

- Using SSA to compute a single trajectory is no more complex than numerical simulation of reaction rate equations.
- Provides a closer approximation of molecular reality for systems with small molecule counts such as genetic circuits.
- Unfortunately, SSA has a substantial computational cost:
  - Must be run many times (1000s) to produce reasonable statistics while simulations of reaction rate equations only run once.
  - Very slow since \( t \) is equal to \( 1/a_0(x) \) and can be very large when any molecule counts become large.
- When molecule counts increase, relative difference between deterministic and stochastic trajectories decrease.
Let $t$ and $x = x_0$.
Evaluate propensity functions $a_j(x)$ at state $x$.
For each $j$, determine the time, $\tau_j$, until the next $R_j$ reaction:
$$\tau_j = t + \frac{1}{a_j(x)} \ln \left( \frac{1}{\tau_j} \right),$$
where each $\tau_j$ is a unit uniform random number.
Store the $\tau_j$ values in an indexed priority queue $Q$.

**Observations**

- First reaction method requires $m$ random variables per simulation!
- Observation: not all propensities change after a reaction.
- Following three steps are taken during every iteration and take a time proportional to the number of reactions, $m$.
  - Update all $m$ propensity functions, $a_j(x)$.
  - Generate $m$ random numbers and next reaction times.
  - Find the smallest reaction time, $\tau_j$.
- Must eliminate each of these performance bottlenecks.

**Gibson/Bruck’s Next Reaction Method**

- Initialize: $t = t_0$ and $x = x_0$.
- Evaluate propensity functions $a_j(x)$ at state $x$.
- For each $j$, determine the time, $\tau_j$, until the next $R_j$ reaction:
  $$\tau_j = t + \frac{1}{a_j(x)} \ln \left( \frac{1}{\tau_j} \right),$$
  where each $\tau_j$ is a unit uniform random number.
- Let $\mu$ be the reaction whose $\tau_j$ is the smallest.
- Let $\tau$ equal $\tau_j$.
- Determine the new state: $t = t + \tau$ and $x = x + \mu$.
- If $t$ is greater than the desired simulation time then halt.
- Record $(x, t)$ and goto step 2.

**Gibson/Bruck’s Improvements**

- $\tau_j$ and $a_j(x)$ stored for use in future iterations.
- $\tau_j$ uses absolute time to make useful for multiple iterations.
- Dependency graph used to indicate relations between reactions.
  - Has vertex for each $R_j$ and edge from $R_j$ to other reaction that has as a reactant either a reactant or a product of $R_j$.
- Reuse every $\tau_j$ except the one for $\tau_m$, renormalizing $\tau_j$ when its propensity has changed as indicated by the dependency graph.
- Indexed priority queue used to organize $a_j(x)$ and $\tau_j$ data to make easy to update and to find smallest entry.
  - An indexed priority queue is a tree structure in which the parent always has a lower $\tau_j$ value than both its children.
  - This means the top node always has the smallest $\tau_j$ value.
  - Can be updated in $O(\log(m))$ time.
Determine value for \( \tau \).

* Tau-leaping gives up exactness to improve simulation speed.

- Many reactions are fired at once in the time interval \([t, t + \tau]\).
- Introduce \( m \) random functions, \( K_j(t; x, t) \), where each returns the number of times that \( R_j \) fires in \([t, t + \tau]\) in state \( X(t) = x \).
- New state after \( \tau \)-leap is:

\[
X(t + \tau) = x + \sum_{j=1}^{m} K_j(t; x, t) v_j.
\]

Discussion

- \( \epsilon \) provides means of trading off accuracy for runtime.
- For large \( \epsilon \), a significant runtime improvement can be achieved at the cost of some accuracy.
- Care has to be taken though as large jumps can cause bad things such as species counts being made negative.
- As \( \epsilon \) is made smaller, tau-leaping gradually reduces to the SSA.
- For a very small \( \epsilon \), not as efficient as SSA as it takes many \( \tau \) leaps that produce no events.
- If \( \tau \) much less than a few multiples of \( 1/a_0(x) \), revert to SSA.
The Chemical Langevin Equation

- If \( \Delta t = \tau \) is small enough that no \( a_j(x) \) changes significantly,
  \[
  X(t + \Delta t) \approx x + \sum_{j=1}^{m} \int_{t}^{t+\Delta t} a_j(x) \, dt \sqrt{a_j(x) \Delta t},
  \]
- If \( \Delta t \) is large enough that there are many firings of each \( R_j \), Poisson can be approximated with a Normal random variable:
  \[
  X(t + \Delta t) \approx x + \sum_{j=1}^{m} N_j(t) \sqrt{a_j(x) \Delta t},
  \]
  where \( N_j(t) \) are \( m \) statistically independent and temporally uncorrelated Normal random variables with mean 0 and variance 1.
- This is simply the reaction rate equation, but it has been derived from stochastic chemical kinetics.
- Chemical Langevin Equation has two parts:
  - A deterministic part that grows linearly with \( a_j(x) \).
  - A stochastic part that grows proportional to \( \sqrt{a_j(x)} \).
  - \( a_j(x) \) grow in direct proportion to the size of the system.
  - Stochastic part scales relative to deterministic part as the inverse square root of the system size.

The Chemical Langevin Equation (cont)

- As size increases, magnitude of fluctuations diminish until chemical Langevin equation can be reduced to:
  \[
  X(t + \Delta t) \approx X(t) + \sum_{j=1}^{m} v_j a_j(X(t)) \sqrt{\Delta t},
  \]
  where \( N_j(t) \) are \( m \) statistically independent and temporally uncorrelated Normal random variables with mean 0 and variance 1.
- This is simply the reaction rate equation, but it has been derived from stochastic chemical kinetics.
- Reaction rate equations valid when system large enough that no propensity changes significantly in \( dt \) and every \( R_j \) fires many times in \( dt \).