Simulation of Chemical Reactions

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

November 8, 2017

Predator-Prey model (Lotka-Volterra system)

A simple Predator-Prey model

- Populations of X prey molecules and Y predator molecules
- Three possible reactions (events)

  $\Rightarrow$ Prey reproduction: $X \rightarrow 2X$

  $\Rightarrow$ Prey consumption to generate predator: $X + Y \rightarrow 2Y$

  $\Rightarrow$ Predator death: $Y \rightarrow \emptyset$

- Each prey reproduces at rate α

  $\Rightarrow$ Population of X preys $\Rightarrow \alpha X =$ rate of first reaction
- Prey individual consumed by predator individual on chance encounter

  $\Rightarrow$ X prey and Y predator $\Rightarrow \beta XY =$ rate of second reaction

  $\Rightarrow$ $\beta =$ Rate of encounters between prey and predator individuals
- Each predator dies off at rate γ

  $\Rightarrow$ Population of Y predators $\Rightarrow \gamma Y =$ rate of third reaction

The Lotka-Volterra equations

- Study population dynamic $\Rightarrow X(t)$ and $Y(t)$ as functions of time t
- Conventional approach: model system as system of differential eqs.
  $\Rightarrow$ Lotka-Volterra (LV) differential equations

- Change in prey $(dX(t)/dt)$ = Prey generation - Prey consumption
- Prey is generated when it reproduces $\Rightarrow$ rate $\alpha X(t)$
- Prey consumed by predators $\Rightarrow$ rate $\beta X(t)Y(t)$

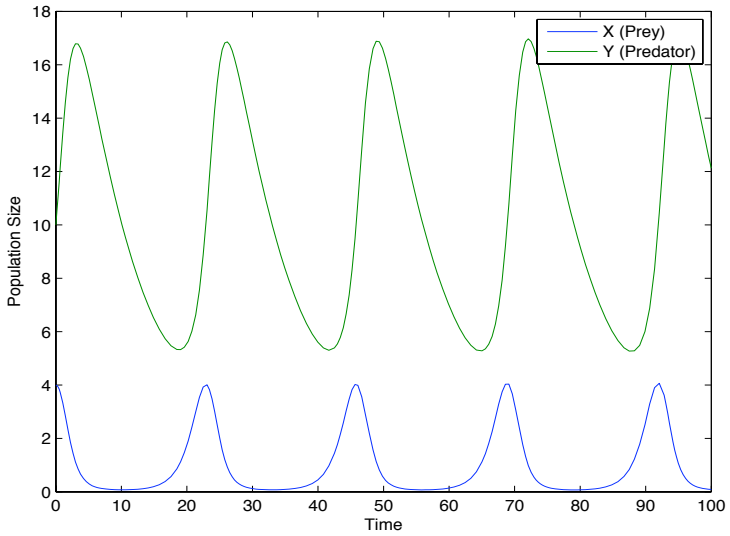
$$\frac{dX(t)}{dt} = \alpha X(t) - \beta X(t)Y(t)$$

- Predator change $(dY(t)/dt)$ = Predator generation - consumption
- Predator is generated when it consumes prey $\Rightarrow$ rate $\beta X(t)Y(t)$
- Predator consumed when it dies off $\Rightarrow$ rate $\gamma Y(t)$

$$\frac{dY(t)}{dt} = \beta X(t)Y(t) - \gamma Y(t)$$

Solution of the LV equations

- LV equations are non-linear but can be solved numerically

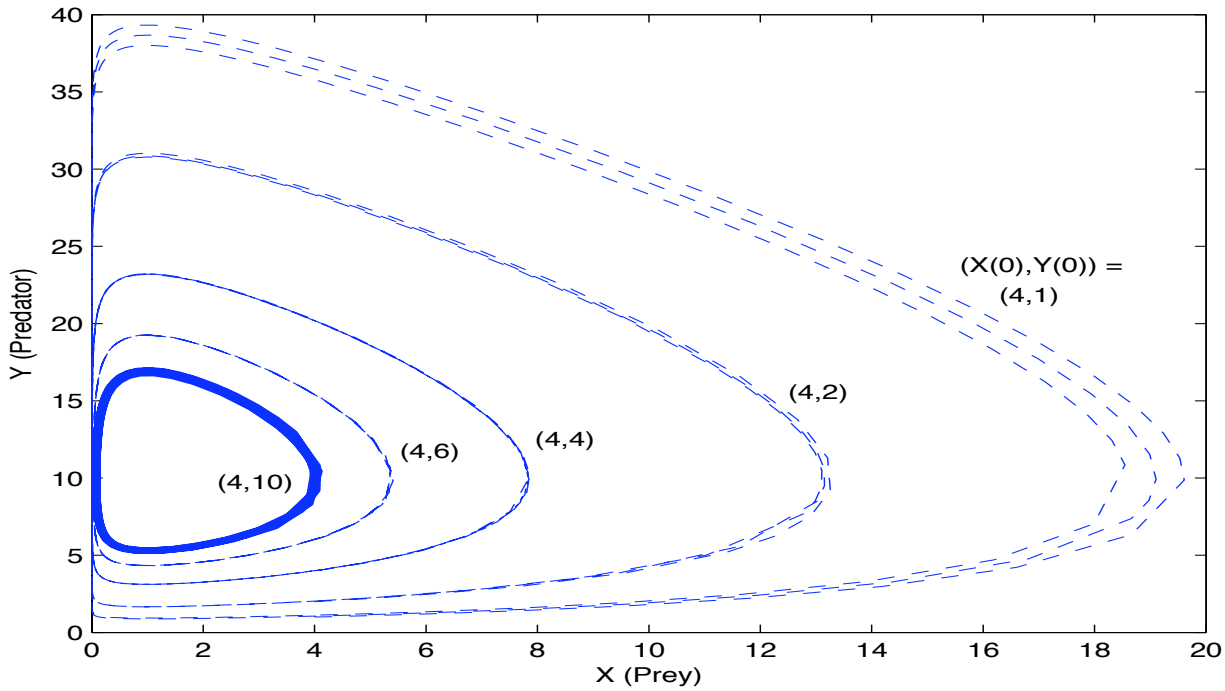


- Prey reproduction rate $\alpha = 1$
- Predator death rate $\gamma = 0.1$
- Predator consumption of prey $\beta = 0.1$
- Initial state $X(0) = 4$ $Y(0) = 10$
- Boom and bust cycles

- Start with prey reproduction $>$ consumption $\Rightarrow$ prey $X(t)$ increases
- Predator production picks up (proportional to $X(t)Y(t)$)
- Predator production $>$ death $\Rightarrow$ predator $Y(t)$ increases
- Eventually prey reproduction $<$ consumption $\Rightarrow$ prey $X(t)$ decreases
- Predator production slows down (proportional to $X(t)Y(t)$)
- Predator production $<$ death $\Rightarrow$ predator $Y(t)$ decreases
- Prey reproduction $>$ consumption (start over)

State space diagram

- State-space diagram $\Rightarrow$ plot $Y(t)$ versus $X(t)$
- System constrained to single orbit given by initial state $X(0)$, $Y(0)$



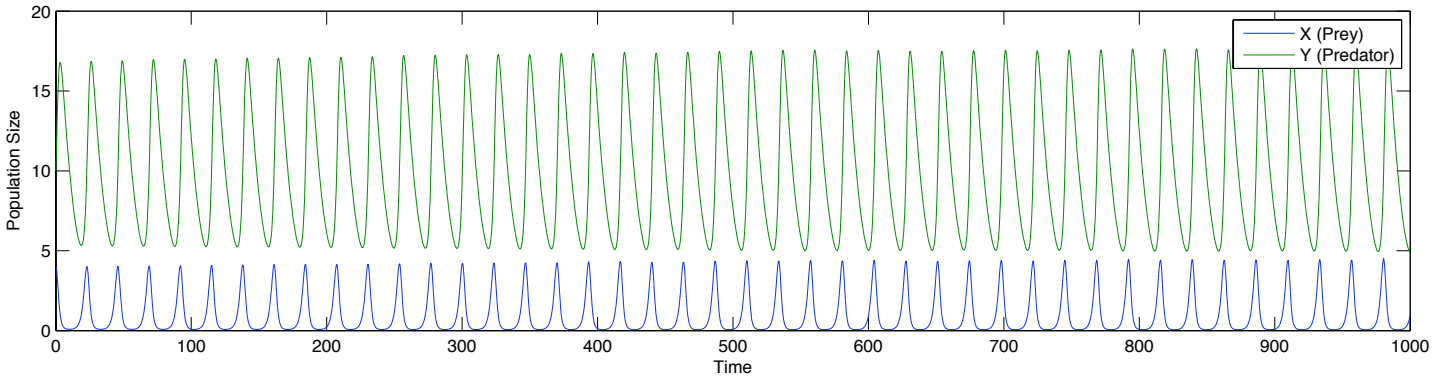
Buildup: Prey increases fast, predator increases slowly (move right and slightly up)

Boom: Predator increases fast depleting prey (move up and left)

Bust: When prey is depleted predator collapses (move down almost straight)

Two observations

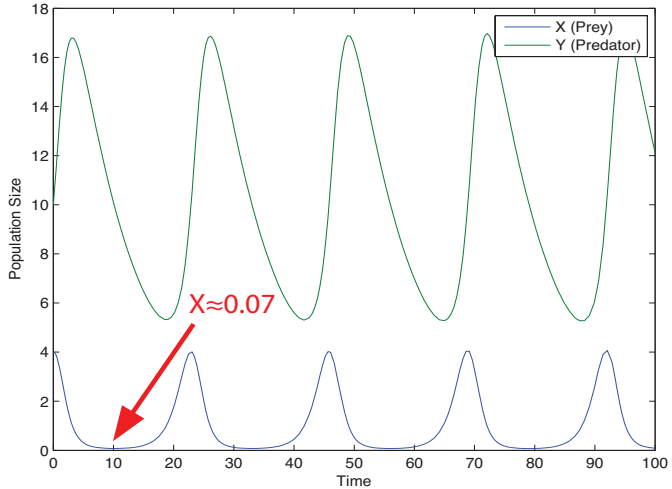
- Too much regularity for a natural system (exact periodicity forever)



- $X(t)$, $Y(t)$ modeled as continuous but actually discrete. Is this a problem?

- If $X(t)$, $Y(t)$ large can interpret as concentrations (molecules/volume)
- Accurate in many cases (millions of molecules)

- If $X(t)$, $Y(t)$ small does not make sense
- Our simulation had 7/100 prey at some point
- There is an extinction event we are missing



Things deterministic model explains (or does not)

- Deterministic model is useful $\Rightarrow$ E.g. boom and bust cycles
  $\Rightarrow$ Important property that the model predicts and explains
- But it does not capture some aspects of the system. E.g.,
  $\Rightarrow$ Non-discrete population sizes (unrealistic fractional molecules)
  $\Rightarrow$ No random variation (unrealistic regularity)
- Possibly missing important phenomena $\Rightarrow$ e.g., extinction
- Shortcomings most pronounced when number of molecules is small
- Important in biochemistry at cellular level ($1 \sim 5$ molecules typical)
- Address these shortcomings through a stochastic model

Stochastic model

- Three possible reactions (events) occurring at rates c_1, c_2 and c_3
  - $\Rightarrow$ Prey reproduction: $X \xrightarrow{c_1} 2X$
  - $\Rightarrow$ Prey consumption to generate predator: $X + Y \xrightarrow{c_2} 2Y$
  - $\Rightarrow$ Predator death: $Y \xrightarrow{c_3} \emptyset$
- Denote as $X(t)$, $Y(t)$ number of molecules by time t
- Can model $X(t)$, $Y(t)$ as continuous time Markov chains (CTMCs)?
- Large population size argument not applicable because we want to model systems with small number of molecules

Stochastic model (continued)

- Consider system with 1 prey molecule x and 1 predator molecule y
- Let $T_2(1,1)$ be the time until x reacts with y
- Since $T_2(1,1)$ is the time until x encounters y and x and y move randomly around it is reasonable to model $T_2(1,1)$ as memoryless

$$\mathrm{P}\left[T_2(1,1) > s+t \,\middle|\, T_2(1,1) > s\right] = \mathrm{P}\left[T_2(1,1) > t\right]$$

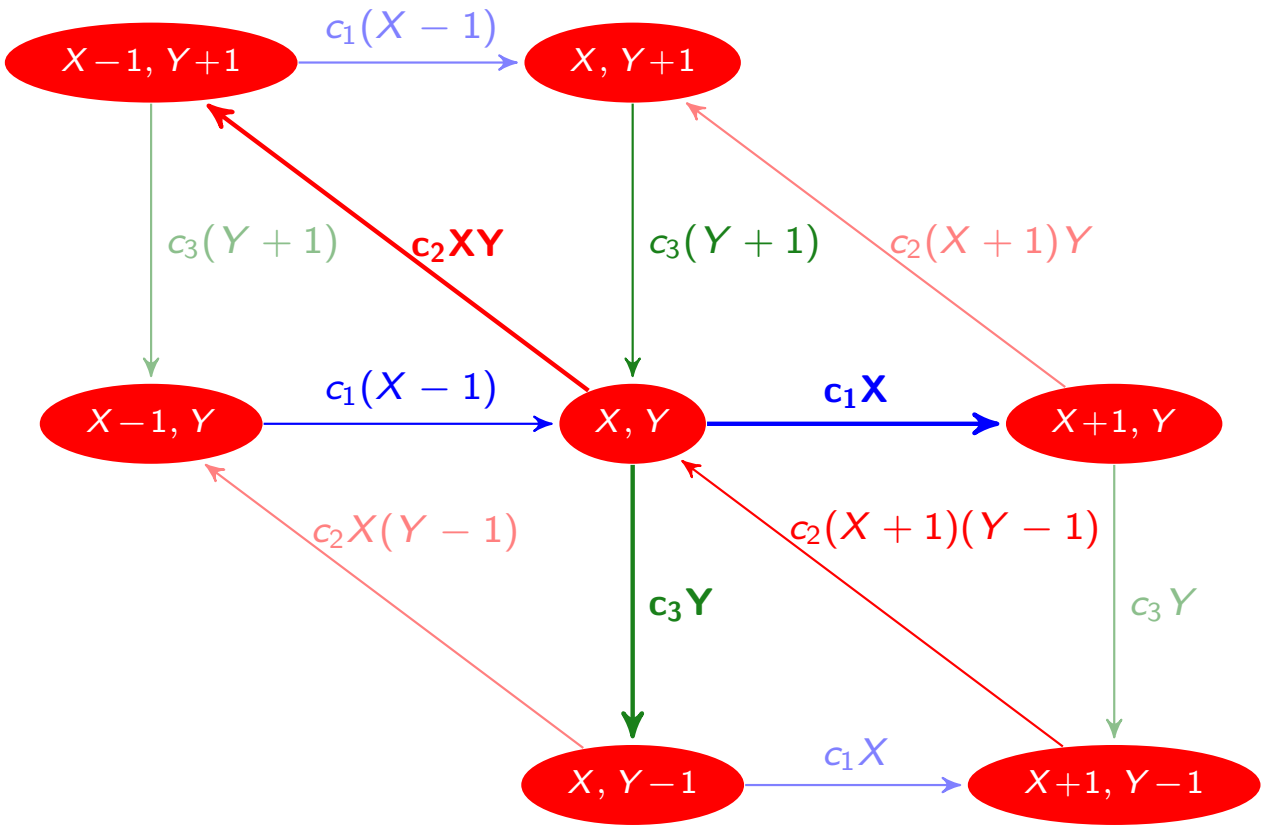
- $T_2(1,1)$ is exponential with parameter c_2
- If there are X prey and Y predator there are XY possible reactions between a specimen of type X and a specimen of type Y
- Let $T_2(X,Y)$ be the time until the first of these reactions occurs
- Min. of exponential RVs is exponential with summed parameters

  $\Rightarrow$ $T_2(X,Y)$ is exponential with parameter c_2XY

- Likewise time $T_1(X)$ until first reaction of type 1 is exponential with parameter c_1X and time $T_3(Y)$ is exponential with parameter c_3Y

CTMC model

- If reaction times are exponential can model as CTMC
- CTMC state is pair (X, Y) with nr. of prey and predator molecules

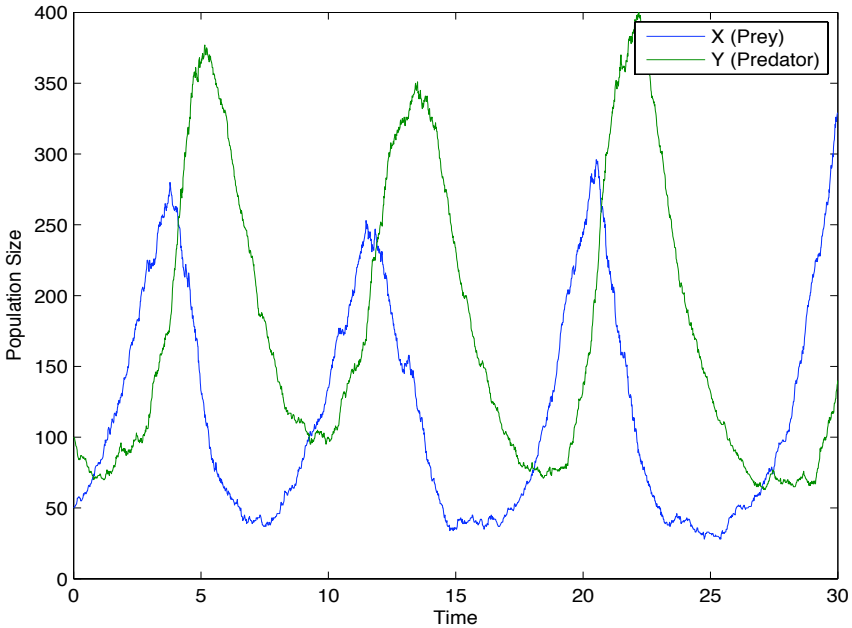


Transition rates

- $(X, Y) \to (X+1, Y)$: Reaction 1 $= c_1X$
- $(X, Y) \to (X-1, Y+1)$: Reaction 2 $= c_2X$
- $(X, Y) \to (X, Y-1)$: Reaction 3 $= c_3X$

Simulation of CTMC model

- Use CTMC model to simulate Predator-prey model
- Initial conditions are $X(0) = 50$ prey and $Y(0) = 100$ predator

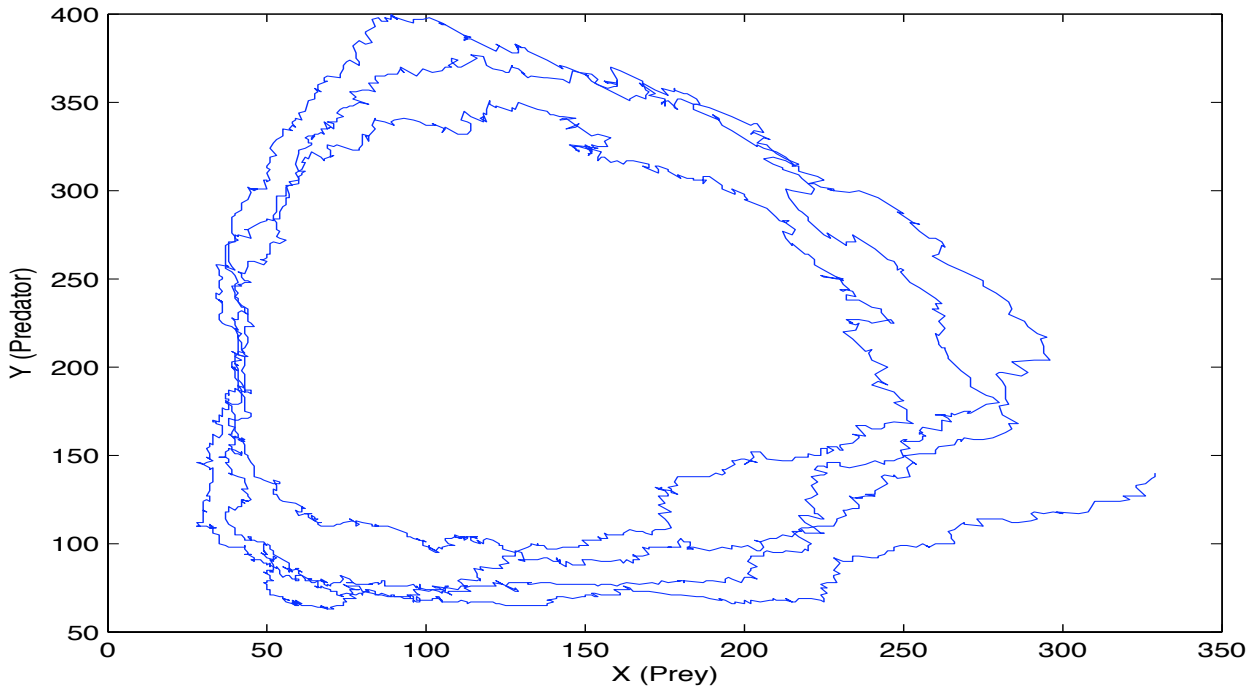


- Prey reproduction rate
  $c_1 = 1$ reactions/second
- Rate of predator consumption of prey
  $c_2 = 0.005$ reactions/second
- Predator death rate
  $c_3 = 0.6$ reactions/second

- Boom and bust cycles are still the dominant feature of the system but random variations are apparent

CTMC model in state space

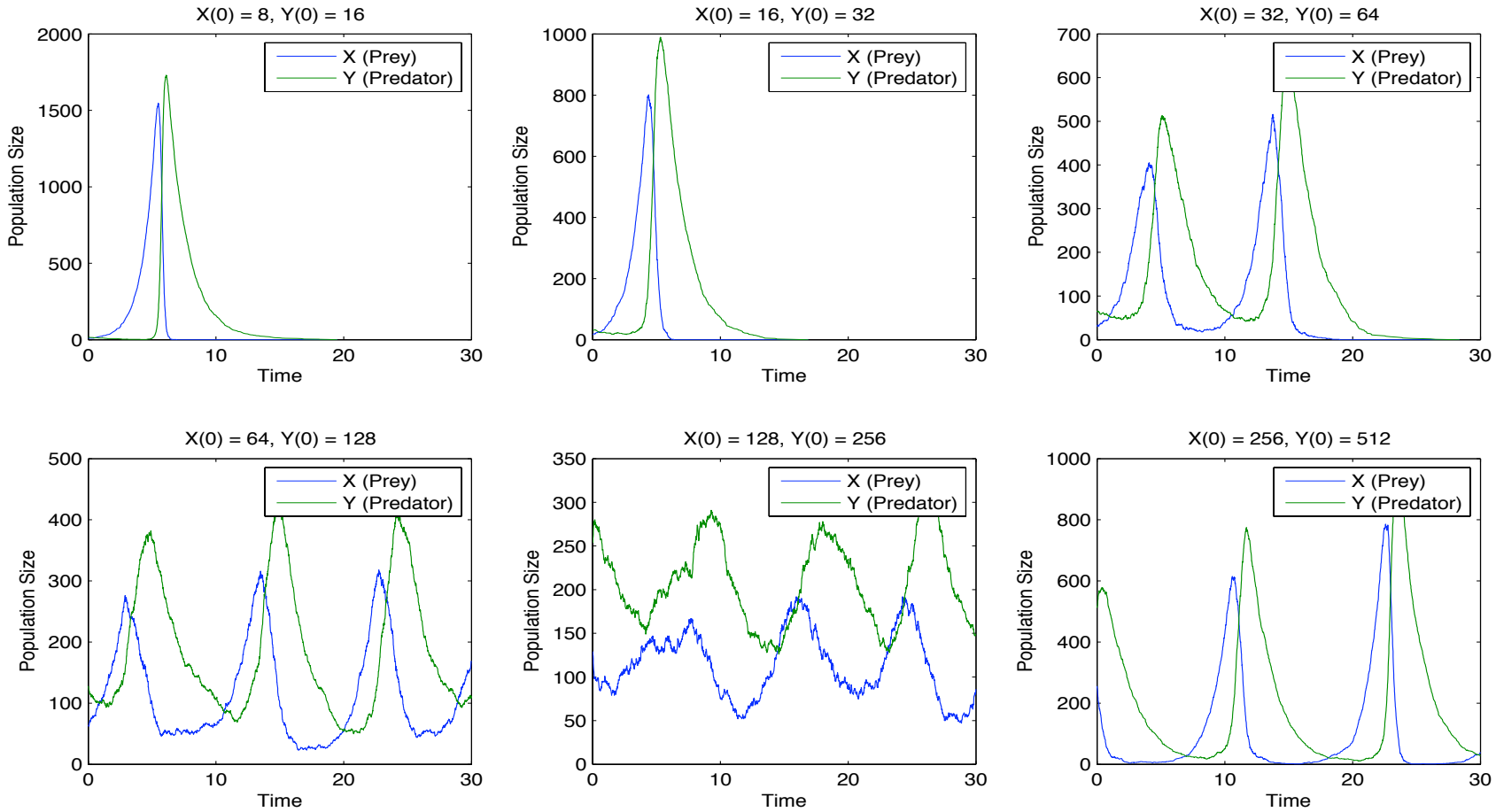
- Plot $Y(t)$ versus $X(t)$ for the CTMC $\Rightarrow$ state space representation



- There is not a single fixed orbit as before
- Can think of this orbit as a perturbed version of deterministic orbit

Effects of different population sizes

- Chance of extinction captured by CTMC model (top plots)



(Notice that Y-axis scales are different)

Conclusions and the road ahead

- Deterministic vs. stochastic modeling
- Deterministic modeling is simpler
  $\Rightarrow$ Captures dominant features (boom & bust cycles)
- Stochastic simulation more complex
  $\Rightarrow$ Less regularity, (all runs are different, state orbit not fixed)
  $\Rightarrow$ Captures effects missed by deterministic solution (extinction)

- Gillespie's algorithm. Forthcoming
- Building a CTMC model for every system of reactions is cumbersome
- Impossible if there are tens or hundreds of types and reactions
- Gillespie's algorithm is just a general way of writing a simulation code for a generic system of chemical reactions

Gillespie's algorithm

Predator-Prey model (Lotka-Volterra system)

Gillespie's algorithm

Dimerization Kinetics

Enzymatic Reactions

Auto-regulatory gene network

Lactose digestion (lac operon)

Simulation of chemical reactions

- Chemical system with m reactant types and n possible reactions
- Reactant quantities change over time as reactions occur
- Nr. of type j reactants at time t denoted as $X_j(t)$
- System's state $\Rightarrow$ vector $\mathbf{X}(t) := [X_1(t), X_2(t), \ldots, X_j(t)]^T$
- To specify i-th reaction $\Rightarrow$ reactants, products and rates

$$R_i : s_{i1}^l X_1 + s_{i2}^l X_2 + \ldots + s_{im}^l X_m \overset{h_i(\mathbf{X})}{\rightarrow} s_{i1}^r X_1 + s_{i2}^r X_2 + \ldots + s_{im}^r X_m$$

- (s_{i1}^l molecules of type 1) $+ \ldots +$ (s_{im}^l molecules of type m) react ... ... to yield (s_{i1}^r of type 1) $+ \ldots +$ (s_{im}^r of type m)
- Rate of reaction $h_i(\mathbf{X})$ depends on number of molecules present
- Let $T_i(\mathbf{X})$ denote the time until the i-th reaction when state is $\mathbf{X}$

Stoichiometry matrices

- Can be more conveniently written using matrices
  - $\Rightarrow$ Define vector of rates $\mathbf{h}(\mathbf{X}) = [h_1(\mathbf{X}), h_2(\mathbf{X}), \ldots, h_n(\mathbf{X})]^T$
  - $\Rightarrow$ Define stoichiometry left matrix $\mathbf{S}^{(l)}$ with elements s_{ij}^l
  - $\Rightarrow$ Define stoichiometry right matrix $\mathbf{S}^{(r)}$ with elements s_{ij}^r
- Write system of chemical reactions as $\Rightarrow \mathbf{S}^{(l)}\mathbf{X} \overset{\mathbf{h}(\mathbf{X})}{\rightarrow} \mathbf{S}^{(r)}\mathbf{X}$

$$\begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ X_m \end{pmatrix} = \mathbf{X} \qquad \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ X_m \end{pmatrix} = \mathbf{X}$$

$$\underbrace{\begin{pmatrix} s_{11}^l & s_{12}^l & \cdot & s_{1m}^l \\ \cdot & \cdot & \cdot & \cdot \\ s_{i1}^l & s_{i2}^l & \cdot & s_{im}^l \\ \cdot & \cdot & \cdot & \cdot \\ s_{n2}^l & s_{n2}^l & \cdot & s_{nm}^l \end{pmatrix}}_{\mathbf{S}^{(l)}} \underbrace{\begin{pmatrix} s_{11}^l X_1 + \ldots + s_{1m}^l X_m \\ \cdot \\ s_{i1}^l X_1 + \ldots + s_{im}^l X_m \\ \cdot \\ s_{n1}^l X_1 + \ldots + s_{nm}^l X_m \end{pmatrix}}_{\mathbf{S}^{(l)}\mathbf{X}} = \underbrace{\begin{pmatrix} s_{11}^r & s_{12}^r & \cdot & s_{1m}^r \\ \cdot & \cdot & \cdot & \cdot \\ s_{i1}^r & s_{i2}^r & \cdot & s_{im}^r \\ \cdot & \cdot & \cdot & \cdot \\ s_{n2}^r & s_{n2}^r & \cdot & s_{nm}^r \end{pmatrix}}_{\mathbf{S}^{(r)}} \underbrace{\begin{pmatrix} s_{11}^r X_1 + \ldots + s_{1m}^r X_m \\ \cdot \\ s_{i1}^r X_1 + \ldots + s_{im}^r X_m \\ \cdot \\ s_{n1}^r X_1 + \ldots + s_{nm}^r X_m \end{pmatrix}}_{\mathbf{S}^{(r)}\mathbf{X}}$$

Example 1: Dimerization kinetics

- Molecule can exist in simple form P and as a dimer D
- Define vector $\mathbf{X} := [P, D]^T$
- Possible reactions are dimerization and dissociation

$$R_1 \text{ (Dimerization): } 2P \overset{h_1(\mathbf{X})}{\rightarrow} D$$

$$R_2 \text{ (Dissociation): } D \overset{h_2(\mathbf{X})}{\rightarrow} 2P$$

- Rates and stoichiometry matrices $\mathbf{S}^{(l)}$ and $\mathbf{S}^{(r)}$ given by

$$\mathbf{S}^{(l)} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{S}^{(r)} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{h}(\mathbf{X}) = \begin{bmatrix} h_1(\mathbf{X}) \\ h_2(\mathbf{X}) \end{bmatrix}$$

- Rewrite equations more compactly as $\Rightarrow \mathbf{S}^{(l)}\mathbf{X} \overset{\mathbf{h}(\mathbf{X})}{\rightarrow} \mathbf{S}^{(r)}\mathbf{X}$

Example 2: Enzymatic reaction

- Substrate S converted to product P. Enzyme E catalyzes conversion
- Converting S into P directly requires significant energy
- Enzyme E reacts with S to form intermediate molecule SE (binding)
- Molecule SE then separates into product P liberating E (conversion)
- This cycle requires less energy than direct conversion
- SE may also separate back into S and E (dissociation)

- Possible reactions are binding, conversion and dissociation, then

$$R_1 \text{ (Binding):} \quad S + E \overset{h_1(\mathbf{X})}{\rightarrow} SE$$

$$R_2 \text{ (Dissociation):} \quad SE \overset{h_2(\mathbf{X})}{\rightarrow} S + E$$

$$R_3 \text{ (Conversion):} \quad SE \overset{h_3(\mathbf{X})}{\rightarrow} E + P$$

Example 2: Enzymatic reaction (continued)

- System state represented by vector $\mathbf{X} := [S, E, SE, P]^T$
- Stoichiometry matrices $\mathbf{S}^{(l)}$ and $\mathbf{S}^{(r)}$ given by

$$\mathbf{S}^{(l)} = \begin{array}{c} \begin{matrix} S & E & SE & P \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} \qquad \mathbf{S}^{(r)} = \begin{array}{c} \begin{matrix} S & E & SE & P \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix}$$

- Reaction rate vector $\mathbf{h}(\mathbf{X}) = [h_1(\mathbf{X}), h_2(\mathbf{X}), h_3(\mathbf{X})]^T$
- Rewrite equations more compactly as $\Rightarrow \mathbf{S}^{(l)}\mathbf{X} \overset{\mathbf{h}(\mathbf{X})}{\rightarrow} \mathbf{S}^{(r)}\mathbf{X}$

Second order reaction

- Consider second order reaction $R_i : X_1 + X_2 \to \ldots$ (two reactants)
- Let $T_i(X_1, X_2)$ be time until R occurs when there are X_1 type 1 and X_2 type 2 molecules
- Have seen that $T_i(X_1, X_2)$ is exponentially distributed with rate

$$h_i(\mathbf{X}) = h_i(X_1, X_2) = c_i X_1 X_2$$

- Constant c_i measures reactivity of X_1 and X_2
- Argument $\Rightarrow$ $T_i(1,1)$ memoryless (depends on chance encounter)
  $\Rightarrow$ Thus $T_i(1,1)$ is exponential with, say, parameter c_i
  $\Rightarrow$ $T_i(X_1, X_2)$ is the minimum of $X_1 X_2$ exponentials
  $\Rightarrow$ $T_i(X_1, X_2)$ exponential with parameter $c_i X_1 X_2$

Second order involving molecules of same type

- Second order reaction with two molecules of same type

$$R_i : X_1 + X_1 \to \dots$$

- Hazard depends on the number of molecules X_1, i.e. $h_i(\mathbf{X}) = h_i(X_1)$
- Reaction does not occur if there is a single molecule
- If there are 2 molecules $T_i(2)$ is exponential with parameter, say, c_i
- For arbitrary X_1 there are $X_1(X_1 - 1)/2$ possible encounters
- Then, $T_i(X_1)$ is exponential with parameter

$$h_i(\mathbf{X}) = h_i(X_1) = c_i X_1(X_1 - 1)/2$$

- $c_i X_1(X_1 - 1)/2$ substantially different from $c_i X_1^2/2$ for small X_1

Zero-th, first and higher order reactions

- Zero-th order reaction $R_i : \emptyset \to X_1$ (spontaneous generation)
- Assume an exponential model with constant rate $h_i = c_i$
- Used to model exogenous factors (and biblical phenomena)

- First order reaction $R_i : X_1 \to \dots$ (decay)
- Exponential with rate $h_i(\mathbf{X}) = h_i(X_1) = c_i X_1$

- Higher order reactions involving more than two reactants
- E.g., third order reaction $R_i : X_1 + X_2 + X_3 \to X_4$
- Time until next R_i reaction exponential. Hazard: $h_i(\mathbf{X}) = c_i X_1 X_2 X_3$
- Reactions of order more than 2 are rare
- Most likely, R_i is encapsulating two second order reactions

$$X_1 + X_2 \to X_5, \qquad X_5 + X_3 \to X_4$$

The hazard function

- All reaction times are exponential RVs $\Rightarrow$ CTMC with state $\mathbf{X}$
- Hazards $h_i(\mathbf{X})$ determine transition rates of CTMC
- Hazards for zero-th, first and second order reactions (for reference)

Order	Reaction		Rate
zero-th	$\emptyset$	$\xrightarrow{c} \cdots$	c
first	X_1	$\xrightarrow{c} \cdots$	cX_1
second	$X_1 + X_2$	$\xrightarrow{c} \cdots$	cX_1X_2
second	$2X_1$	$\xrightarrow{c} \cdots$	$cX_1(X_1 - 1)/2$

- Probability of reaction R_i happening in infinitesimal time ϵ is

$$\mathrm{P}\left[T_i(\mathbf{X}) < \epsilon\right] = h_i(\mathbf{X})\epsilon + o(\epsilon)$$

- That's why the name hazard

State transition for given reaction

- State is $\mathbf{X}(t) = \mathbf{X}$. Reaction R_i occurs. Next state $\mathbf{X}(t + dt) = \mathbf{Y}$?
- Number of reactants per type =
  $= i$-th row of left stoichiometry matrix $\mathbf{s}_i^{(l)} = [s_{i1}^l, s_{i2}^l, \ldots, s_{im}^l]^T$

$$s_{i1}^l X_1 + s_{i2}^l X_2 + \ldots + s_{im}^l X_m \overset{h_i(\mathbf{X})}{\rightarrow} \ldots$$

- Number of products per type =
  $= i$-th row of right stoichiometry matrix $\mathbf{s}_i^{(r)} = [s_{i1}^r, s_{i2}^r, \ldots, s_{im}^r]^T$

$$\ldots \overset{h_i(\mathbf{X})}{\rightarrow} s_{i1}^r X_1 + s_{i2}^r X_2 + \ldots + s_{im}^r X_m$$

- X decreases by nr. of reactants and increases by nr. of products
- Next sate is $\Rightarrow \mathbf{Y} = \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}$ (upon reaction R_i)

Transition rates and probabilities

- $q(\mathbf{X}, \mathbf{Y})$ = transition rate from state $\mathbf{X}$ to state $\mathbf{Y}$. Given by

$$q\left(\mathbf{X}, \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}\right) = h_i(\mathbf{X}), \quad i = 1, \ldots, n$$

- Transition from state $\mathbf{X}$ to $\mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}$ when reaction R_i occurs
- $\nu(\mathbf{X})$ = Transition rate out of $\mathbf{X}$ into any state (any reaction occurs)

$$\nu(\mathbf{X}) = \sum_{i=1}^{n} q\left(\mathbf{X}, \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}\right) = \sum_{i=1}^{n} h_i(\mathbf{X})$$

- $P(\mathbf{X}, \mathbf{Y})$ = Prob. of going into $\mathbf{Y}$ given transition out of $\mathbf{X}$ occurs

$$P\left(\mathbf{X}, \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}\right) = \frac{q\left(\mathbf{X}, \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}\right)}{\nu(\mathbf{X})} = \frac{h_i(\mathbf{X})}{\nu(\mathbf{X})}$$

- Probability that i-th reaction occurs given that a reaction occurred

Gillespie's algorithm

Gillespie's algorithm = Simulation of CTMC

Input: Stoichiometry matrices $\mathbf{S}^{(l)}$ and $\mathbf{S}^{(r)}$. Initial state $\mathbf{X}(0)$

Output: Molecule numbers as a function of time $\mathbf{X}(t)$

(1) Initialize time and CTMC's state $t = 0,\ \mathbf{X} = \mathbf{X}(0)$

(2) Calculate all hazards $\Rightarrow h_i(\mathbf{X})$

(3) Calculate transition rate $\Rightarrow \nu(\mathbf{X}) = \sum_{i=1}^{n} h_i(\mathbf{X})$

(4) Draw random time of next reaction $\Delta t \sim \text{Exp}\big(\nu(\mathbf{X})\big)$

(5) Advance time to $t = t + \Delta t$

(6) Draw reaction at time $t + \Delta t$ $\Rightarrow R_i$ drawn with prob. $h_i(\mathbf{X})/\nu(\mathbf{X})$

(7) Update state vector to account for this reaction $\Rightarrow \mathbf{X} - \mathbf{s}_i^{(l)} + \mathbf{s}_i^{(r)}$

(8) Repeat from (2)

Dimerization Kinetics

Predator-Prey model (Lotka-Volterra system)

Gillespie's algorithm

Dimerization Kinetics

Enzymatic Reactions

Auto-regulatory gene network

Lactose digestion (lac operon)

Dimerization

- Dimerization occurs when two like molecules join together
- Many proteins (P) will form dimers (D)
- Dimerization may be rare in relative terms, but significant in absolute terms at high concentration. For this reason plays important role in auto-regulation of protein production
- Possible reactions are dimerization and dissociation

$$R_1 \text{ (Dimerization): } 2P \xrightarrow{c_1} D$$

$$R_2 \text{ (Dissociation): } D \xrightarrow{c_2} 2P$$

- Dimerization rare and dimers unstable $\Rightarrow c_2 \gg c_1$
- Stoichiometry matrices $\mathbf{S}^{(l)}$ and $\mathbf{S}^{(r)}$ given by

$$\mathbf{S}^{(l)} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{S}^{(r)} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix},$$

- Rate of reaction 1 is $h_1(\mathbf{X}) = c_1 P(P-1)/2$. Reaction 2 is $h_2(\mathbf{X}) = c_2 D$

Gillespie's algorithm for dimerization kinetics

(1) Initialize time and CTMC's state $t = 0$, $P = P(0)$, $D = D(0)$

(2) Calculate hazards $\Rightarrow h_1(\mathbf{X}) = c_1 P(P-1)/2$,
$\Rightarrow h_2(\mathbf{X}) = c_2 D$

(3) Calculate transition rate $\Rightarrow \nu(\mathbf{X}) = c_1 P(P-1)/2 + c_2 D$

(4) Draw random time of next reaction

$$\Delta t \sim \exp\left(\nu(\mathbf{X})\right) = \exp\left(c_1 P(P-1)/2 + c_2 D\right)$$

(5) Advance time to $t = t + \Delta t$

(6) Draw reaction at time $t + \Delta t$

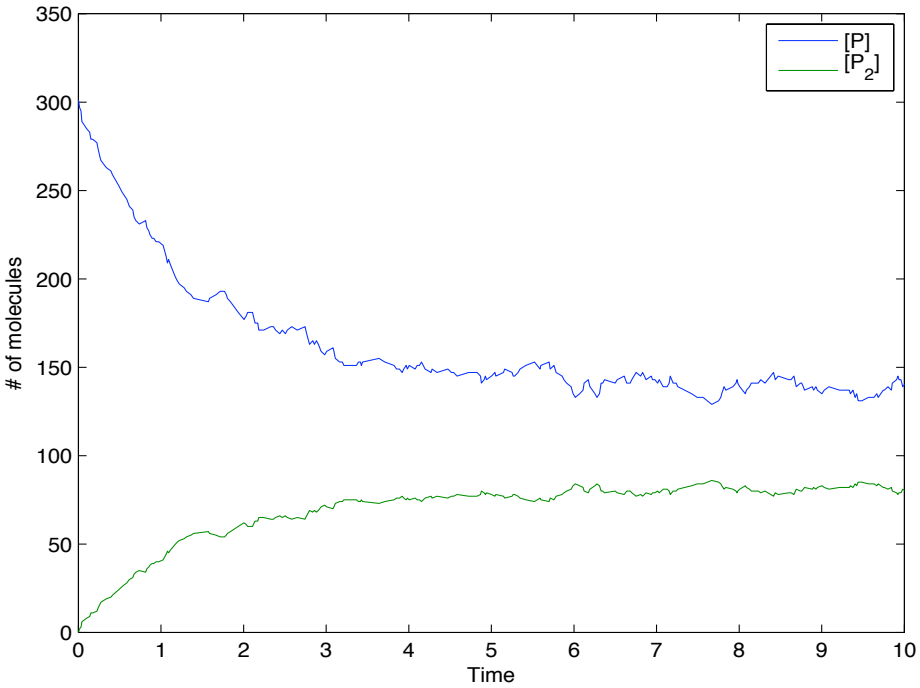
$$\begin{aligned} \mathrm{P}\left[\text{Dimerization:}\right] &= c_1 P(P-1)/2/\nu(\mathbf{X}) \\ \mathrm{P}\left[\text{Dissociation:}\right] &= c_2 D/\nu(\mathbf{X}) \end{aligned}$$

(7) Update state vector $\Rightarrow$ Dimerization: $P = P - 2$, $D = D + 1$
$\Rightarrow$ Dissociation: $P = P + 2$, $D = D - 1$

(8) Repeat from (2)

Stochastic simulation of dimerization kinetics

- Run of Gillespie's algorithm for dimerization kinetics
- Initial condition $P(0) = 301$, $D(0) = 0$ (protein only)



- Dimerization hazard

$$c_1 = 1.66 \times 10^{-3} \frac{\text{reactions}}{\text{sec./molecule}^2}$$

- Dissociation hazards

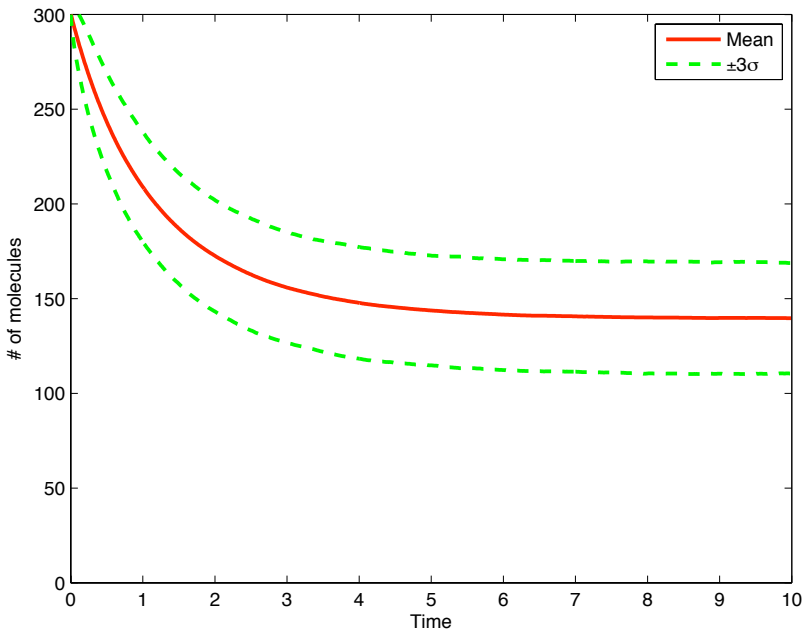
$$c_2 = 0.2 \times 10^{-3} \frac{\text{reactions}}{\text{sec./molecule}}$$

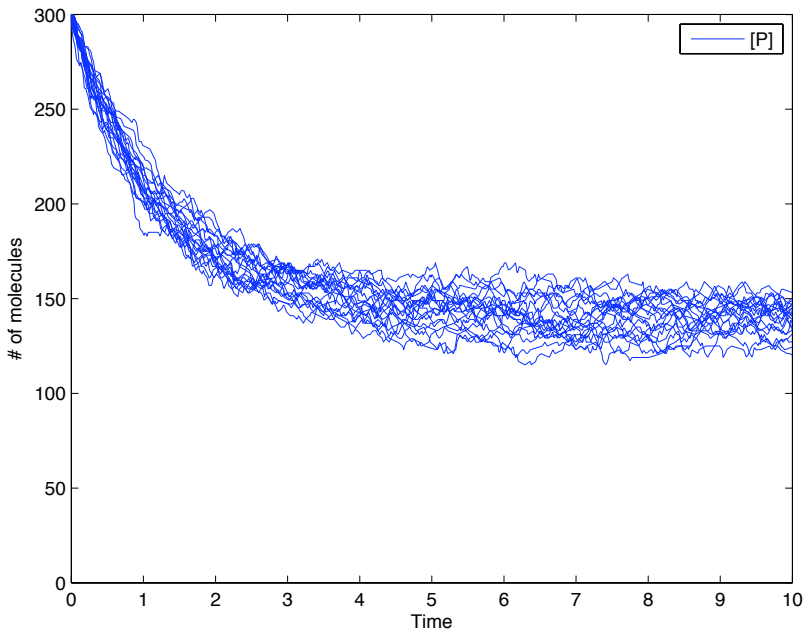
- $\mathbf{c} = [c_1, c_2]^T = [1.66 \times 10^{-3}, 0.2]^T$

- P and D "stabilize" at point where dimerization and dissociation become equally likely

Information that can be obtained from simulations

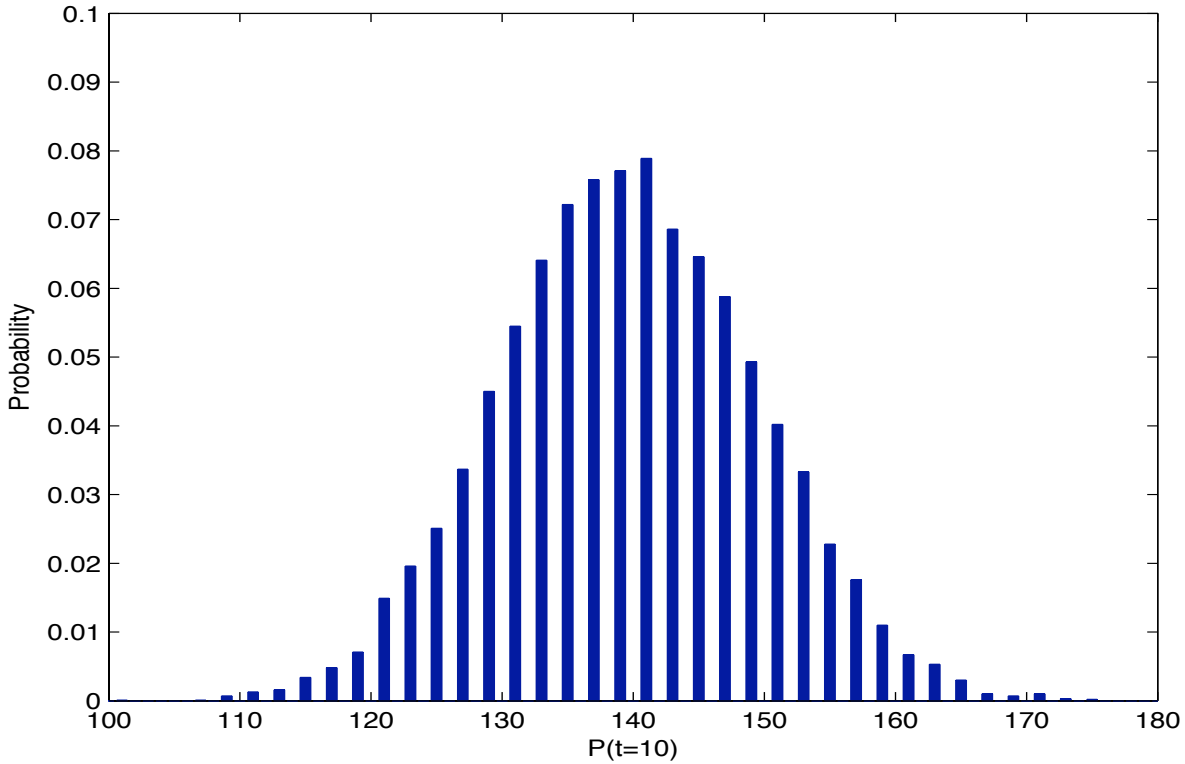
- E.g., consider nr. of protein molecules P ($P(t) + 2D(t)$ is constant)
- Mean and standard deviation of P versus time?
- Right graph $\Rightarrow$ mean and ± 3(standard deviations) over 10^4 trials
- Left graph shows 20 trials
  - Vary around mean path but stay within ± 3-standard deviations





Steady-state probability distribution

- Time $t = 10$ seconds $\Rightarrow$ approximate PMF over 10^4 trials
- Can use ergodicity instead



- Bell-shaped. Only odd values of P are possible
- Runs are all odd or all even depending on initial condition

Enzymatic reactions

Predator-Prey model (Lotka-Volterra system)

Gillespie's algorithm

Dimerization Kinetics

Enzymatic Reactions

Auto-regulatory gene network

Lactose digestion (lac operon)

Enzymes

- Substrate S converted into product P by action of enzyme E
- Intermediate product SE generated by combination of E and S
- SE later separates into product P liberating the enzyme E
- SE may also dissociate into S and E
- Enzymes can act as catalysts for reactions that would otherwise rarely or never take place

- Possible reactions are binding, dissociation and conversion

$$
\begin{aligned}
R_1 \text{ (Binding):} \quad & S + E \xrightarrow{c_1} SE \\
R_2 \text{ (Dissociation):} \quad & SE \xrightarrow{c_2} S + E \\
R_3 \text{ (Conversion):} \quad & SE \xrightarrow{c_3} P + E
\end{aligned}
$$

- Dissociation typically not significant because $c_2 \ll c_3$

Enzymatic reactions (continued)

- Stoichiometry matrices $\mathbf{S}^{(l)}$ and $\mathbf{S}^{(r)}$ given by

$$\mathbf{S}^{(l)} = \begin{array}{c} \begin{matrix} S & E & SE & P \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} \qquad \mathbf{S}^{(r)} = \begin{array}{c} \begin{matrix} S & E & SE & P \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix}$$

- Reaction rates are
  - $\Rightarrow$ Reaction R_1 (Binding): $h_1(\mathbf{X}) = c_1 S \times E$,
  - $\Rightarrow$ Reaction R_2 (Dissociation): $h_2(\mathbf{X}) = c_2 SE$
  - $\Rightarrow$ Reaction R_3 (Conversion): $h_3(\mathbf{X}) = c_3 SE$

Gillespie's algorithm for enzymatic reactions

(1) Initialization: $t = 0$, $S = S(0)$, $E = E(0)$, $SE = SE(0)$, $P = P(0)$

(2) Calculate hazards $\Rightarrow h_1(\mathbf{X}) = c_1 S \times E$,
$\Rightarrow h_2(\mathbf{X}) = c_2 SE$
$\Rightarrow h_3(\mathbf{X}) = c_3 SE$

(3) Calculate transition rate $\Rightarrow \nu(\mathbf{X}) = c_1 S \times E + c_2 SE + c_3 SE$

(4) Draw random time of next reaction

$$\Delta t \sim \exp\left(\nu(\mathbf{X})\right) = \exp\left(c_1 S \times E + c_2 SE + c_3 SE\right)$$

(5) Advance time to $t = t + \Delta t$

(6) Draw reaction at time $t + \Delta t$

$$\begin{aligned} \mathrm{P}\,[\text{Binding:}] &= c_1 S \times E / \nu(\mathbf{X}) \\ \mathrm{P}\,[\text{Dissociation:}] &= c_2 SE / \nu(\mathbf{X}) \\ \mathrm{P}\,[\text{Conversion:}] &= c_3 SE / \nu(\mathbf{X}) \end{aligned}$$

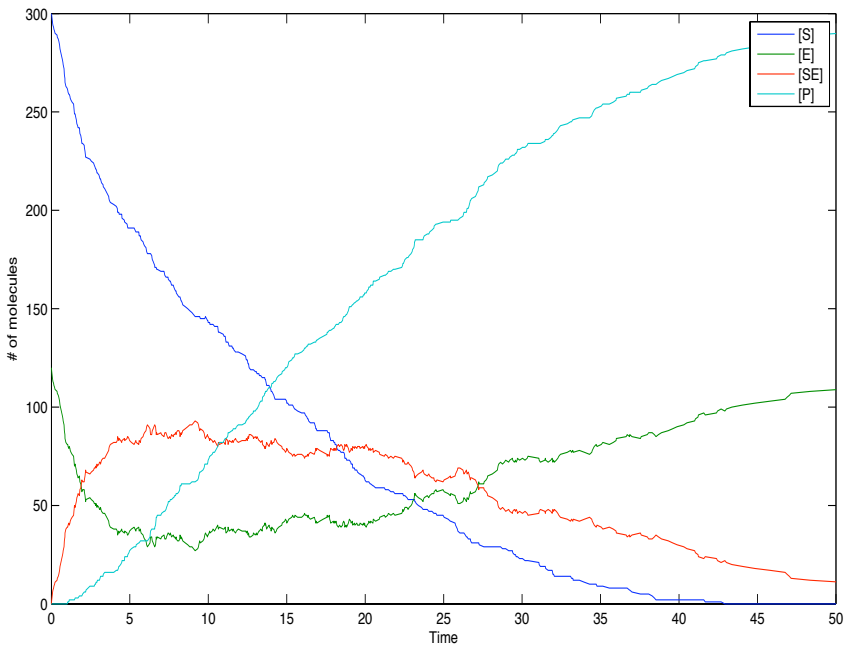
(7) Update state vector $\Rightarrow$ Binding: $S = S - 1$, $E = E - 1$, $SE = SE + 1$
$\Rightarrow$ Dissociation: $S = S + 1$, $E = E + 1$, $SE = SE - 1$
$\Rightarrow$ Conversion: $P = P + 1$, $E = E + 1$, $SE = SE - 1$

(8) Repeat from (2)

Stochastic simulation of enzymatic reactions

- Run of Gillespie's algorithm for enzymatic reactions
- Initialize with only substrate and enzyme present

$$S(0) = 301,\ E(0) = 120,\ SE(0) = 0,\ P(0) = 0$$



- Binding hazard

$$c_1 = 1.66 \times 10^{-3} \frac{\text{reactions}}{\text{sec./molecule}^2}$$

- Dissociation hazard

$$c_2 = 10^{-4} \frac{\text{reactions}}{\text{sec./molecule}}$$

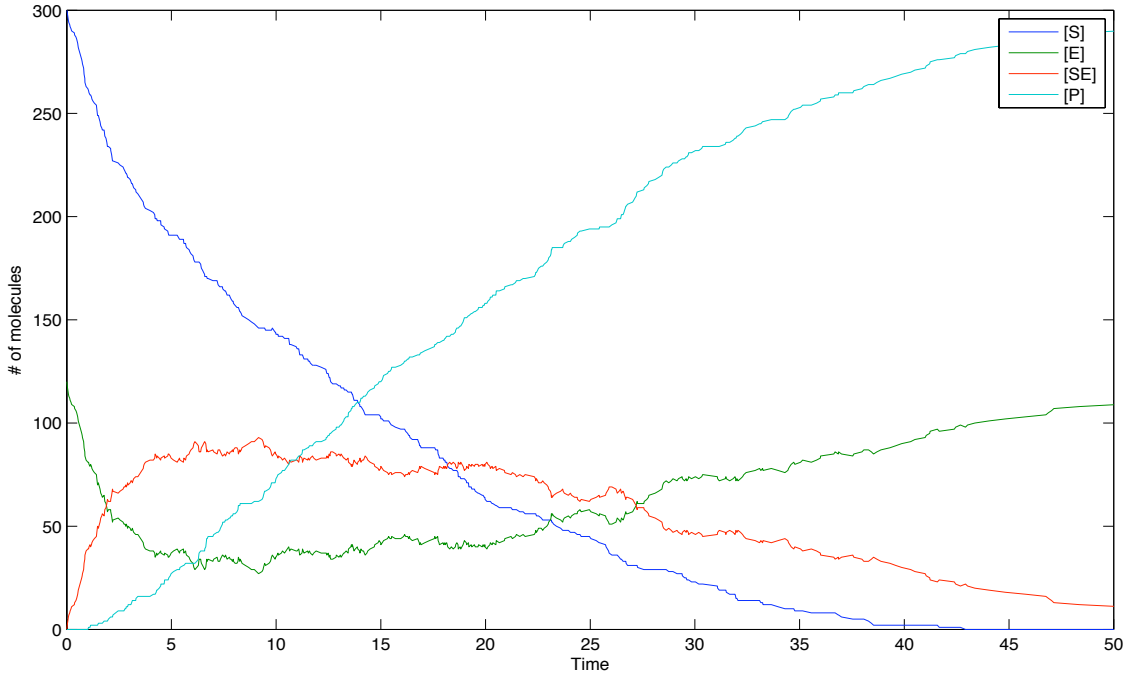
- Conversion hazard

$$c_3 = 0.1 \frac{\text{reactions}}{\text{sec./molecule}}$$

- $\mathbf{c} = [c_1, c_2, c_3]^T = [1.66 \times 10^{-3}, 10^{-4}, 0.1]^T$

Stochastic simulation (continued)

- At the beginning substrate and enzyme numbers decline as they bind to each other to form intermediate product SE
- Intermediate product separates into final product P liberating enzyme E
- By $t = 50$ seconds substrate is completely converted into product and enzymes are free. There is no intermediate product either



Auto-regulatory gene network

Predator-Prey model (Lotka-Volterra system)

Gillespie's algorithm

Dimerization Kinetics

Enzymatic Reactions

Auto-regulatory gene network

Lactose digestion (lac operon)

Auto-regulation of protein production

- Simplified model of protein production in prokaryotes
- "Instructions" for creating protein (P) "encoded" in gene (G)
- To produce protein, gene G is first transcribed into mRNA (R)
- This mRNA is passed on to a ribosome to "assemble" the protein

- Protein production usually triggered by external stimuli
- How is it halted?

  $\Rightarrow$ Negative feedback loops called auto-regulatory networks

- As protein numbers increase, so does presence of a byproduct,
  - E.g., a protein dimer (D)
- Byproducts show affinity to bind to the gene blocking transcription
- Halting transcription slows/halts protein production

Reactions of the auto-regulatory network

- Protein production consists of transcription and assembly

$$\text{Transcription: } G \xrightarrow{c_1} G + R$$

$$\text{Assembly: } R \xrightarrow{c_2} R + P$$

- Dimer is generated as a byproduct of protein production

$$\text{Dimerization: } 2P \xrightarrow{c_3} D$$

$$\text{Dissociation: } D \xrightarrow{c_4} 2P$$

- Dimer binds to mRNA blocking transcription. Blocked gene may be "liberated"

$$\text{Repression: } G + D \xrightarrow{c_5} GD$$

$$\text{Liberation: } GD \xrightarrow{c_6} G + D$$

- Protein and mRNA eventually degrade (mRNA degradation common)

$$\text{mRNA degradation: } R \xrightarrow{c_7} \emptyset$$

$$\text{Protein degradation: } P \xrightarrow{c_8} \emptyset$$

Approach to modeling

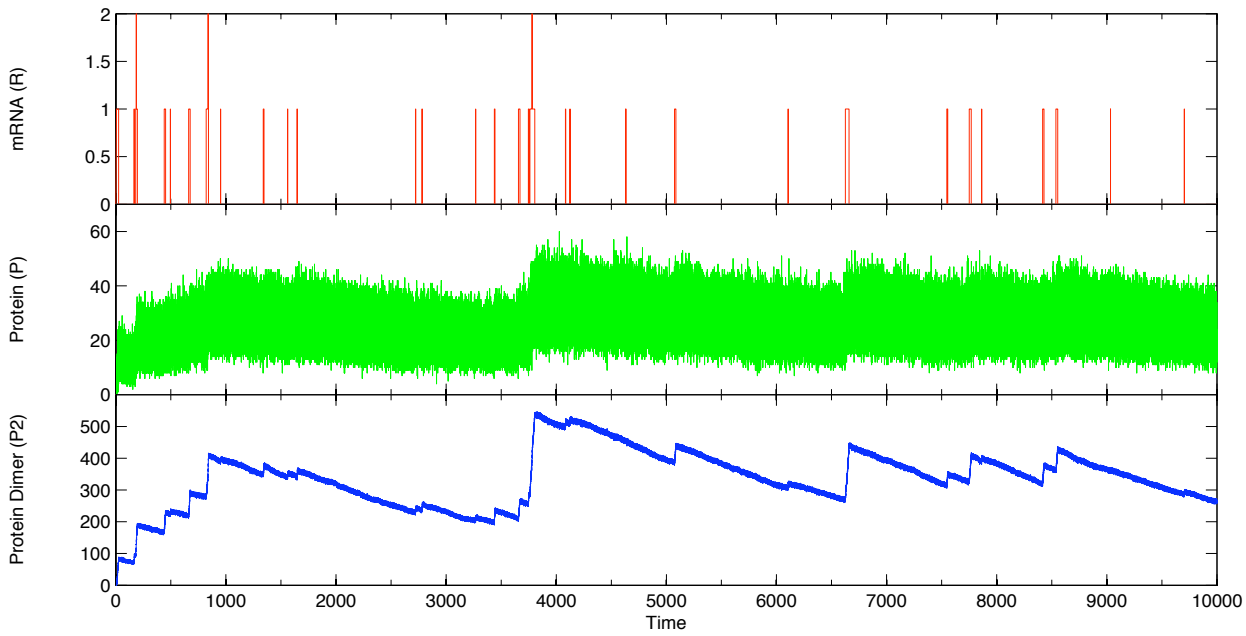
- We will use rate constants

$$\mathbb{C} = \begin{bmatrix} c_1 \text{ (transcription)} & = 0.01 \\ c_2 \text{ (assembly)} & = 10 \\ c_3 \text{ (dimerisation)} & = 1 \\ c_4 \text{ (dissociation)} & = 1 \\ c_5 \text{ (repression)} & = 1 \\ c_6 \text{ (reverse repression)} & = 10 \\ c_7 \text{ (mRNA degradation)} & = 0.1 \\ c_8 \text{ (protein degradation)} & = 0.01 \end{bmatrix}$$

- $\overset{\text{gene}}{G(0)} = 10, \quad \overset{\text{repressed gene}}{P_2G(0)} = \overset{\text{mRNA}}{R(0)} = \overset{\text{protein}}{P(0)} = \overset{\text{dimer}}{P_2(0)} = 0$
- Because of the very small numbers of molecules involved, a continuous deterministic approach would not provide accurate results.

The stochastic solution

- Stochastic simulation. Protein dimer and mRNA numbers shown
- mRNA numbers are very small (0, 1 or 2)



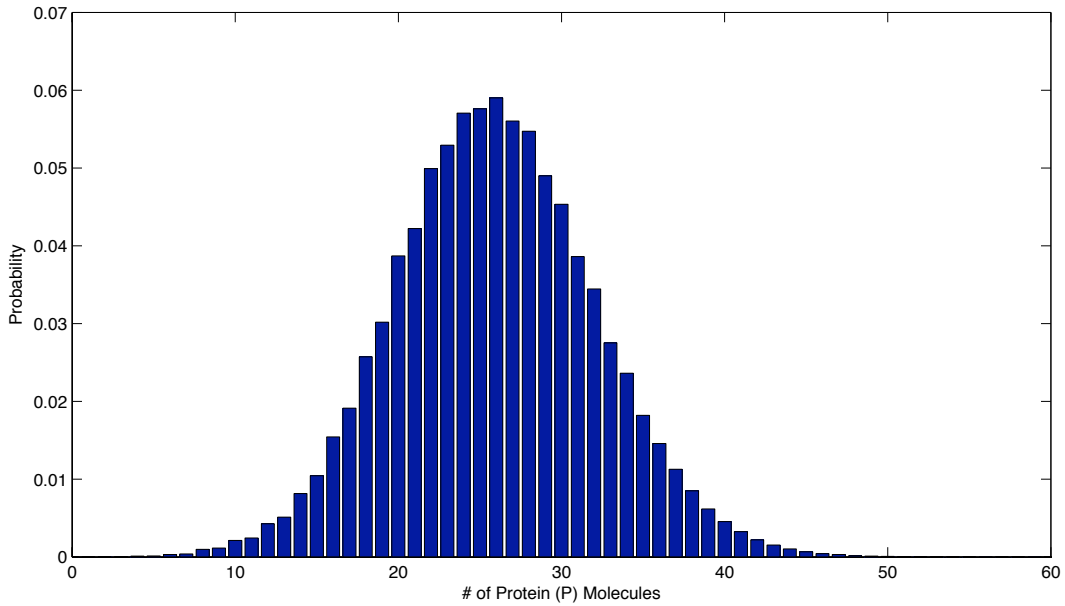
- Increase in protein & dimer triggered by mRNA transcription events
- Transcription events spread out when protein nrs. are large
- Transcription events occur more rapidly when there is less protein

Notes

- Because there are a very small number of genes, the stochastic nature of the number of mRNA molecules transcribed is very clearly evident.
- Even though there are larger numbers of P_2, their numbers are affected directly by the mRNA transcription events, so stochasticity still dominates.

Observing the distribution

- At steady-state, we find the following PMF for the number of protein molecules (over 10,000 trials, using the property of ergodicity):



- Notice that the distribution is very evenly centered around 25, showing successful auto-regulation.

Lactose digestion (lac operon)

Predator-Prey model (Lotka-Volterra system)

Gillespie's algorithm

Dimerization Kinetics

Enzymatic Reactions

Auto-regulatory gene network

Lactose digestion (lac operon)

Auto-regulation of protein production

- Simplified model of protein production in prokaryotes
- "Instructions" for creating proteins "encoded" in genes
- To produce proteins, genes are first transcribed into mRNA
- This mRNA is passed on to a ribosome to "assemble" the protein

- Protein production not immutable. How does it changes over time?
- Auto regulatory gene networks
  - $\Rightarrow$ Production triggered by external stimuli
  - $\Rightarrow$ Halted by negative feedback loops through protein byproducts

- E.g. Production of β-galactosidase to digest glucose
  - $\Rightarrow$ Lac-operon (lac for lactose, operon=set of interacting genes)

Glucose, Lactose and β-galactosidase

- Glucose (G) and lactose (L) are variations of sugars
- Cells use glucose for energy but can reduce lactose to glucose
- Lactose reduced to glucose by enzyme β-galactosidase (βG)

$$\text{Lactose digestion:} \quad L + \beta G \xrightarrow{c_1} G + \beta G$$

$$\text{Glucose consumption:} \quad G \xrightarrow{c_2} \emptyset$$

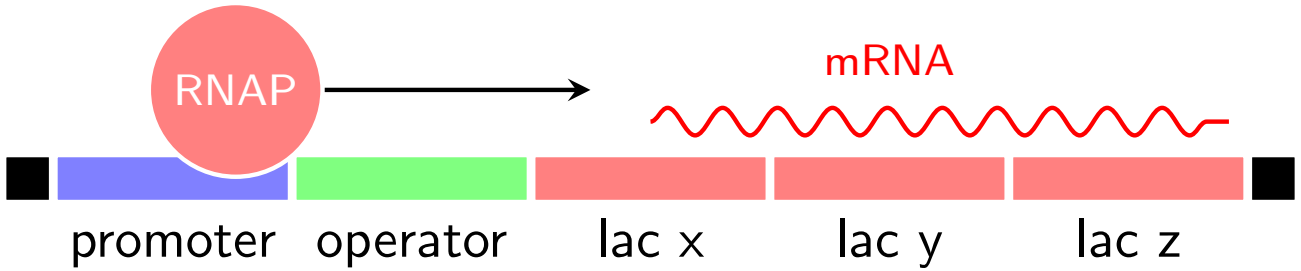
- Did not model enzymatic reaction (compare with earlier example)
- Rate of lactose digestion $c_1 L \times (\beta G)$. Glucose consumption $c_2 G$

- Producing β-galactosidase is not always necessary
- Production necessary only when lactose is present and glucose is not

Lac-operon, normal state

- Lac-operon consists of three adjacent genes
- Promoter, operator and β-galactosidase code (three types in fact)
- Lac-operon has three possible states, regular, activated and repressed

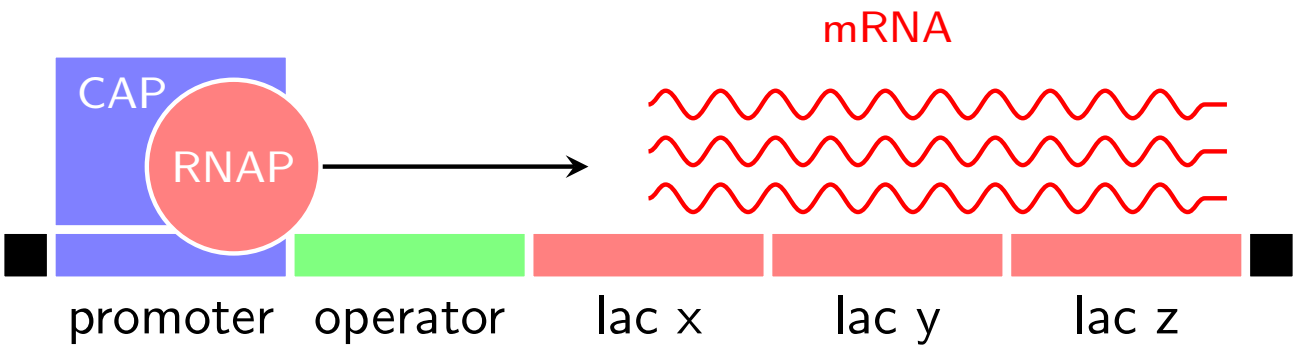
- In normal state (Op) transcription proceeds at a small rate c_3
- The promoter is a binding place for RNA polymerase (RNAP)
- RNAP binds to promoter to initiate gene transcription into mRNA



- Model reaction as $\Rightarrow$ Regular transcription: $Op \xrightarrow{c_3} Op + mRNA$

Lac-operon in activated state

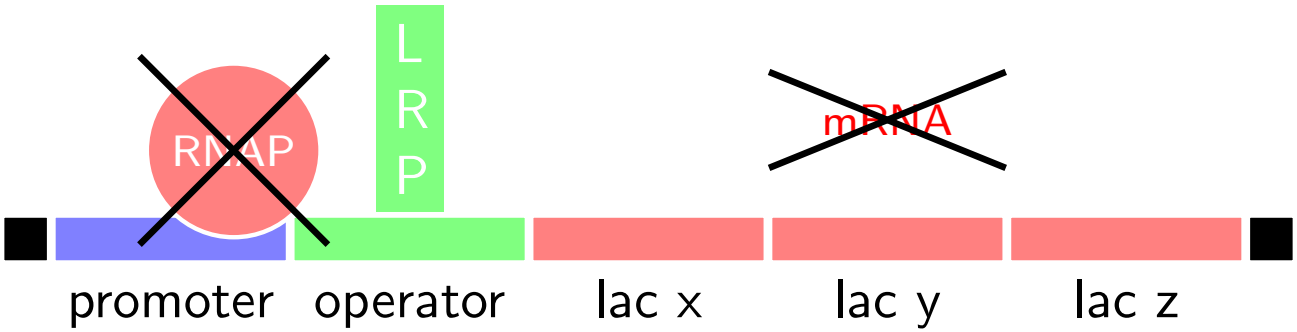
- Operon activated (AOp) by catabolite activator protein (CAP)
- CAP binds upstream of the promoter altering DNA's geometry
- Thereby facilitating (promoting) binding of RNAP to promoter
- Hence yielding a faster rate of transcription $c_4 \gg c_3$



- Model reaction as $\Rightarrow$ Activated transcription: $AOp \xrightarrow{c_4} AOp + mRNA$

Lac-operon in repressed state

- Operon repressed (ROp) by lactose repressor protein protein (LRP)
- LRP encoded by gene adjacent to lac operon, is always expressed and has great affinity with the operator
- If LRP binds to operator it interferes with RNAP–promoter binding
- Without RNAP, there is no (or minimal) transcription
- Hence yielding a very slow rate of transcription $c_5 \ll c_3 \ll c_4$

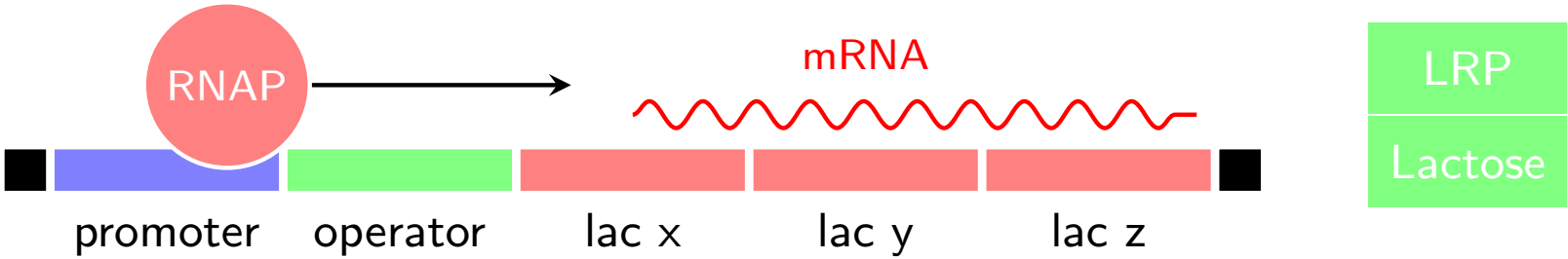


- Model reaction as $\Rightarrow$ Repressed transcription: $ROp \xrightarrow{c_5} ROp + mRNA$

Repression control

- If there is no lactose (L) present lac operon is in repressed state
- When lactose is present it combines with LRP
- Thereby preventing repression of lac operon. Lac operon in regular state

  $\Rightarrow$ Small (but not minimal) rate of β-galactosidase production



- We model this with the following reactions

$$\begin{aligned}
&\text{Operon repression:} && LRP + Op \xrightarrow{c_6} ROp \\
&\text{Operon liberation:} && ROp \xrightarrow{c_7} LRP + Op \\
&\text{Repressor neutralization:} && LRP + L \xrightarrow{c_8} LRPL \\
&\text{Repressor dissociation:} && LRPL \xrightarrow{c_9} LRP + L
\end{aligned}$$

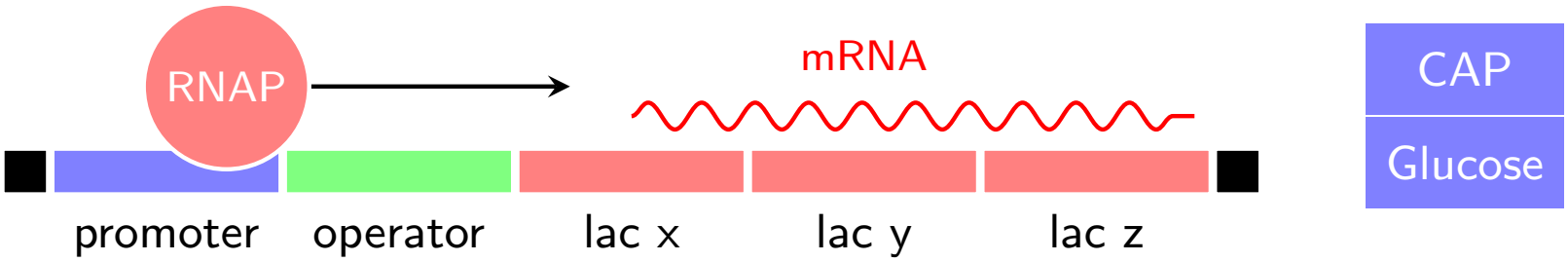
Activation control

- Prevalence of CAP inversely proportional to glucose levels
- This involves a complex set of reactions in itself
- For a preliminary model the following reactions suffice

$$\begin{aligned}
&\text{Operon activation:} && CAP + Op \xrightarrow{c_{10}} AOp \\
&\text{Operon deactivation:} && AOp \xrightarrow{c_{11}} CAP + Op \\
&\text{CAP neutralization:} && CAP + G \xrightarrow{c_{12}} CAPG \\
&\text{CAP dissociation:} && CAPG \xrightarrow{c_{13}} CAP + G
\end{aligned}$$

- If glucose is present, CAP is bound to glucose
- Thereby preventing activation of lac operon

  $\Rightarrow$ Small rate of β-galactosidase production



Glucose, lactose and lac-operon states

- High lactose and high glucose (glucose preferred)
  - CAP bound to glucose and LRP bound to lactose
  - Operon in regular state, low production of β-galactosidase
- High lactose and low glucose (lactose only option)
  - CAP bound upstream of promoter and LRP bound to lactose
  - Operon in activated state, high production of β-galactosidase
- High glucose and low lactose (glucose dominant and preferred)
  - CAP bound to glucose and LRP bound to operator
  - Operon in repressed state, minimal production of β-galactosidase
- Low glucose and low lactose (no energy source available)
  - CAP bound upstream of promoter and LRP bound to operator
  - Repression dominates, minimal production of β-galactosidase
- β-galactosidase produced in significant quantities only with high lactose and low glucose concentrations

β-galactosidase assembly and decays

- To complete model we add reactions to account for
  - $\Rightarrow$ Assembly of β-galactosidase (βG) enzyme
  - $\Rightarrow$ *mRNA* and βG decay

Protein synthesis:	$mRNA \xrightarrow{c_{14}} mRNA + \beta G$
mRNA decay:	$mRNA \xrightarrow{c_{15}} \emptyset$
βgalactosidase decay:	$\beta G \xrightarrow{c_{16}} \emptyset$

Reactions modeling digestion of lactose

- Model of auto-regulatory gene network for digestion of lactose
- Rates in reactions/minute/molecule or reactions/minute/molecule^2

Lactose digestion:	$L + \beta G$	$\xrightarrow{c_1} G + \beta G$	$c_1 = 1$
Glucose consumption:	G	$\xrightarrow{c_2} \emptyset$	$c_2 = 0.1$
Regular transcription:	Op	$\xrightarrow{c_3} Op + mRNA$	$c_3 = 0.01$
Activated transcription:	AOp	$\xrightarrow{c_4} AOp + mRNA$	$c_4 = 0.1$
Repressed transcription:	ROp	$\xrightarrow{c_5} ROp + mRNA$	$c_5 = 0.001$
Operon repression:	$LRP + Op$	$\xrightarrow{c_6} ROp$	$c_6 = 1$
Operon liberation:	ROp	$\xrightarrow{c_7} LRP + Op$	$c_7 = 1$

- Compare rates c_3-c_5 for lac operon in different states

Reactions modeling digestion of lactose (continued)

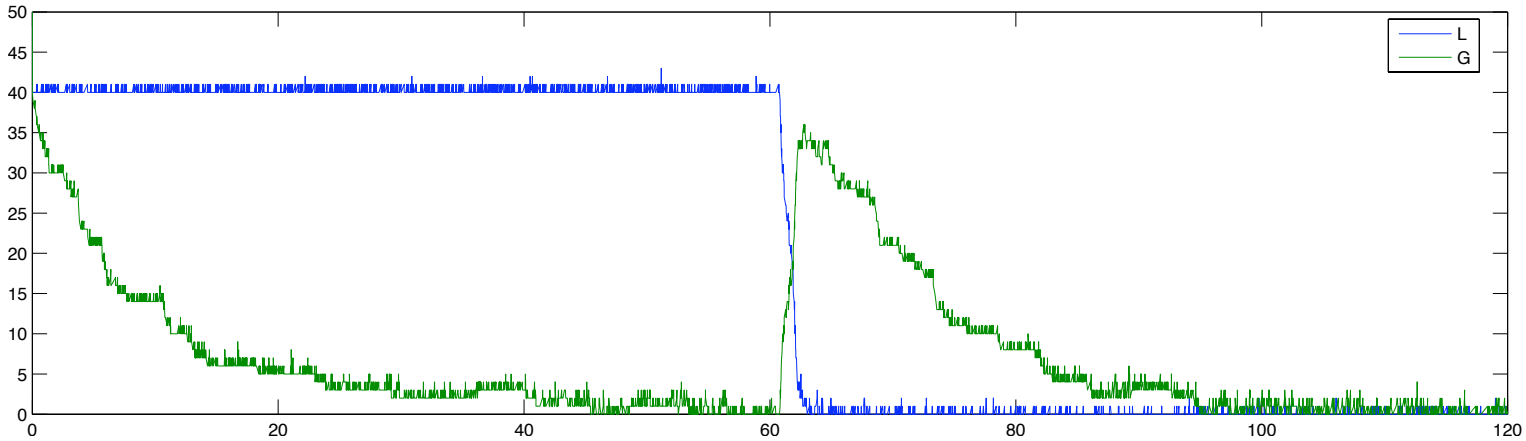
- Model of auto-regulatory gene network for digestion of lactose
- Rates in reactions/minute/molecule or reactions/minute/molecule^2

Repressor neutralization:	$LRP + L$	$\xrightarrow{c_8} LRPL$	$c_8 = 10$
Repressor dissociation:	$LRPL$	$\xrightarrow{c_9} LRP + L$	$c_9 = 1$
Operon activation:	$CAP + Op$	$\xrightarrow{c_{10}} AOp$	$c_{10} = 1$
Operon deactivation:	AOp	$\xrightarrow{c_{11}} CAP + Op$	$c_{11} = 1$
CAP neutralization:	$CAP + G$	$\xrightarrow{c_{12}} CAPG$	$c_{12} = 10$
CAP dissociation:	$CAPG$	$\xrightarrow{c_{13}} CAP + G$	$c_{13} = 1$
Protein synthesis:	$mRNA$	$\xrightarrow{c_{14}} mRNA + \beta G$	$c_{14} = 1$
mRNA decay:	$mRNA$	$\xrightarrow{c_{15}} \emptyset$	$c_{15} = 1$
βgalactosidase decay:	βG	$\xrightarrow{c_{16}} \emptyset$	$c_{16} = 0.1$

- Notice that LRP and CAP neutralization are fast (rates c_8 and c_{12})

Stochastic simulation: diauxie pattern

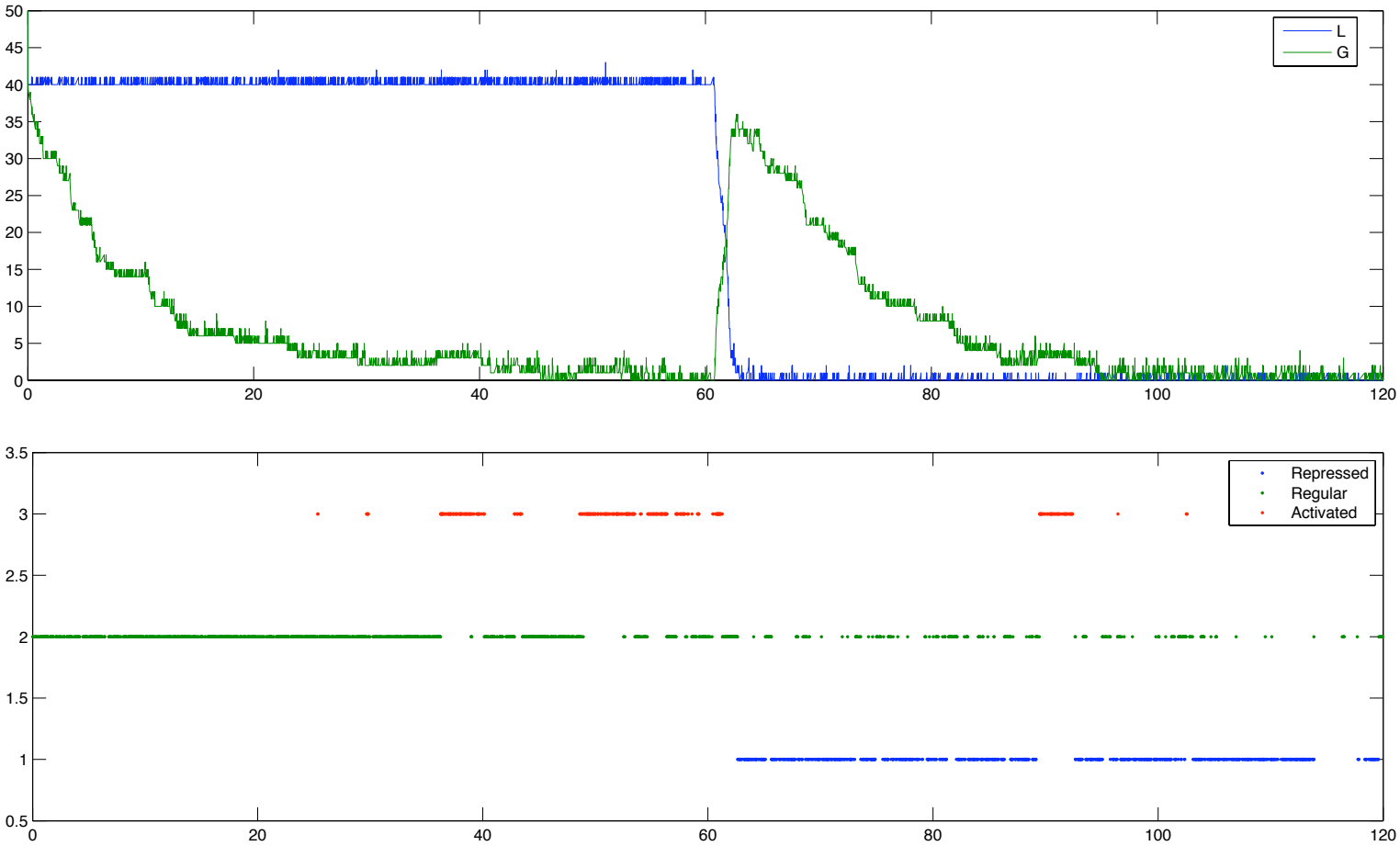
- Initial state $\Rightarrow L = 50,\ G = 50,\ CAP = 10,\ LRP = 10$
- Only 1 operon in regular state



- Sugars (glucose and lactose) consumed sequentially
  - $\Rightarrow$ Glucose is consumed first
  - $\Rightarrow$ After glucose is depleted, lactose converted to glucose
  - $\Rightarrow$ After conversion, newly generated glucose is also consumed
- Yields two growth spurts = diauxie pattern

Operon state and diauxie pattern

- Conversion occurs with operon in activated state



mRNA transcription & β-Galactosidase synthesis

- Operon activation $\Rightarrow$ mRNA transcription $\Rightarrow$ β-Galactosidase synthesis $\Rightarrow$ lactose digestion