

Michaelis-Menten kinetics: explain and derive.

 Set concentration of enzyme abiding to $E_0 = E + E_B$.
 Assume quasi-equilibrium for reduction of substrate (complex forming).
 The system is saturated with the complex.
 Set $K = k_1/(k_2+k_3)$
 Set $V_{\max} = E_0 \cdot k_3$

What's a Hill equation?

 In biochemistry and pharmacology, the binding of a ligand to a macromolecule is often enhanced if there are already other ligands present on the same macromolecule (this is known as cooperative binding). The Hill coefficient provides a way to quantify this effect.

It describes the fraction of the macromolecule saturated by ligand as a function of the ligand concentration; it is used in determining the degree of cooperativeness of the ligand binding to the enzyme or receptor.
 (Ligand = funktionell grupp -- molekyl, atom, etc.)

Has the form $1/((K_A/L)^n + 1) = L^n/(K_D + L^n)$.

How does it apply to transcriptional activation or repression?

 Activation: $S^n/(K_S + S^n)$
 Repression: $1/(K_S + S^n)$

Higher $n \rightarrow$ more abrupt transition (in transcription factor), and thereby determines whether the system is "on" or "off" given a certain amount of TF.

How do you extend these equations to gene regulation by multiple transcription factors?

 SHEA-ACKERS. Partition sums, probability of on/off.
 ASK CARL.

Write down equations for a gene that's activated by A and repressed by B & can you come up with more than one solution?

 ASK CARL.

What are advantages and disadvantages to simplifying equations? (example: the dimensionless equations for a bistable switch)

 Advantages: Fewer free variables when simulating. Ease-of-tweak.
 Can reduce highly complex systems to few significant parts.

Drawbacks: Loss of proper interpretation of mechanisms inherent to the system.
 Also no rigorous interpretation of outcome, as equations are of arbitrary (no) dimension.
 Loss of detail -- might overgloss important, external aspects when modelling.

The model can (probably) be justified by comparison to actual data.

More?

When should you use stochastic simulations?

- 1. When you want to capture the stochasticity of a system.
 2. In investigating stability and the likes.

Describe the Gillespie algorithm.

- 1. Draw two random numbers.
 2. Increment time logarithmically.
 3. Weigh reactions according to their equational significance.
 4. Normalize.
 5. Split up reactions on probability line.

6. Let random number 2 define what reaction will happen.
7. Adjust system accordingly.
8. Repeat.

Why don't we always use stochastic simulations?

If many reactions are present, the system can:

1. take a long, long time to simulate.
2. produce code that is not benign (too hard-coded/conditional).
3. break system (foxes die out, and system thus requires a rescaling -- interpretation lost in translation).
4. mustn't necessarily give other/better results than deterministic case
5. often requires a deterministic implementation for comparison. Extra work.

Gillespie also assumes a well-stirred system.

Can you say anything about alternatives to the basic Gillespie algorithm?

Modifications that don't recompute all probabilities if they haven't changed, which is good for bigger, more complex systems.

1. Next reaction method

Published 2000. This is an improvement over the first reaction method where the unused reaction times are reused. To make the sampling of reactions more efficient, an indexed priority queue is used to store the reaction times. On the other hand, to make the recomputation of propensities more efficient, a dependency graph is used. This dependency graph tells which reaction propensities to update after a particular reaction has fired.

2. Optimised and sorting direct methods

Published 2004 and 2005. These methods sort the cumulative array to reduce the average search depth of the algorithm. The former runs a presimulation to estimate the firing frequency of reactions, whereas the latter sorts the cumulative array on-the-fly.

3. Logarithmic direct method

Published in 2006. This is a binary search on the cumulative array, thus reducing the worst-case time complexity of reaction sampling to $O(\log M)$.

4. Partial-propensity methods

Published in 2009, 2010, and 2011 (Ramaswamy 2009, 2010, 2011). Use factored-out, partial reaction propensities to reduce the computational cost to scale with the number of species in the network, rather than the (larger) number of reactions. Four variants exist.

How would you model passive transport between cells?

Diffusion.

Q: Is there a difference between transport within cells and between cells.

A: Units!

Diffusion: $\text{length}^2/\text{time}$

Membrane permeability: $\text{length}/\text{time}$

Also: statistical physics approach. Probabilities of moving between states. Chemiosmotic transport theory.

How fast is diffusion?

 $t \sim l^2/2D$ (Fick's law)

Cell size $\sim 10 \mu\text{m}$

Travel within cell: $\sim 3 \text{ s}$

Travel a distance of 1 m: $\sim 16 \text{ years(!)}$

How can active transport be modelled?

Enzyme reaction. Many biochemical reactions are mediated by enzymes.

When between cells, remember proper indexing.

Relation between cell A and cell B?

What is chemical potential?

Describes (measures?) availability of reactant. Is defined as proportional to the change in entropy (or free energy) as the particle species number changes.

Special cases:

Dilute solution

Chemical equilibrium ($\mu_a = \mu_b$)

Can be combined with other potentials, e.g. Nernst.

What are network motifs?

An occurrence in the interaction of elements in a network that statistically are more significant than pure random interactions. Example: transcriptional interaction network. Also feedforward loops, coherent ff-loops, incoherent ff-loops.

What dynamics are associated with some common network motifs?

See slides (Network motifs dynamics).

What is a morphogen?

A morphogen is a substance governing the pattern of tissue development in the process of morphogenesis, and the positions of the various specialized cell types within a tissue. More specifically, a morphogen is a signaling molecule that acts directly on cells to produce specific cellular responses depending on its local concentration.

Can for example signal "Hey! Come here. Be a feather!"

What are reaction--diffusion systems? Can you give an example?

Systems that are keen on pattern forming. Usually includes a local, short-range part and a semi-global, long-range part. Example: activator--inhibitor system, activator--substrate.

"Reaction;diffusion systems are mathematical models which explain how the concentration of one or more substances distributed in space changes under the influence of two processes: local chemical reactions in which the substances are transformed into each other, and diffusion which causes the substances to spread out over a surface in space."

What complications arise if cells aren't all the same size?

What are you actually modeling? Concentrations or particle numbers? Do all cells produce the same amount of molecules? Are the equilibrium concentrations the same? When adding growth -- what happens to the number of particles, or the concentration?

Can you say something about cell division?

How to model this is not currently very well known; different ideas exist. Henrik uses methods for cell division rules that reach back to the 19th Bcentury:

1. Hofmeister (1863) perpendicular to main axis of growth
2. Sachs (1878) perpendicular to old walls
3. Errara (1888) shortest path into two equal volumes (Henrik's favourite)

And about mechanics in spatial model?

Stress can explain direction of growth. See Henrik's notes.

What is the Lotka-Volterra model? What assumptions is it based on?

The Lotka;Volterra equations, also known as the predator;prey equations, are a pair of first-order, non-linear, differential equations frequently used to describe the dyna

mics
of biological systems in which two species interact, one as a predator and the other as prey.

Assumptions:

1. Unlimited food for prey (x)
2. Predators limited only by prey
3. Predators have unlimited appetite
4. Individuals are ageless

Can you design a more realistic model for a simple predator-prey system?

Limited food.
Limited growth rate.
Stochasticity.
Intra-species competition.
More details, e.g. age, health, sex

Consider an epidemiological model such as the SIR model. Can you interpret the parameters?

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta I S}{N} \\ \frac{dI}{dt} &= \frac{\beta I S}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

S -- Susceptible
 I -- Infected
 R -- Recovered
 N -- total population
 β -- parameter of infectivity
 γ -- rate of recovery

Some notes on the infection rate (force of infection):

Note that in the above model the function

$F = \beta \frac{I}{N}$,

models the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals, so that it is called the force of infection. However, for large classes of communicable diseases it is more realistic to consider a force of infection that does not depend on the absolute number of infectious subjects, but on their fraction (with respect to the total constant population N):

$F = \beta \frac{I}{N}$

Capasso and, afterwards, other authors have proposed nonlinear forces of infection to model more realistically the contagion process.

How would you extend the SIR model to include death? loss of resistance?

Extra category. People who die come from the infected group. Dead people stay dead (often).

The SIR model with vital dynamics and constant population:

Considering a population characterized by a death rate μ and birth rate λ , and where a communicable disease is spreading. The model with mass-action transmission is

$$\begin{aligned} \frac{dS}{dt} &= \lambda - \mu S - \beta I S \\ \frac{dI}{dt} &= \beta I S - (\gamma + \mu) I \\ \frac{dR}{dt} &= \gamma I - \mu R \end{aligned}$$

Can also have loss-of-resistance: SIS, SIRS (temporal resistance, i.e. fraction of R back to S).

How could these population models be extended to include spatial aspects?

Use nodes and inter-nodal travel, grids, etc. Diffusion?

What problems arise when simulating spatial models stochastically?

Instability.

Number of reactions. Hard to keep track of what happens when and where? Could potentially be solved with alternative Gillespie methods.
Concentration--Particle number issues?

ASK CARL?

What types of optimization methods can be used for fitting models to data?

Local optimization: Gradient descent and variations

Global and stochastic optimization: Simulated annealing, genetic algorithms

Briefly describe the gradient based methods.

Steepest descent: The search direction is $\mathbf{p}_k = -\nabla f(\mathbf{x}_{k+1})$
Every new step is always perpendicular to the old one.

Conjugate descent:

We start with the first direction given by the gradient. We choose the new direction so as we move along it the gradient in the old direction remains 0.
$$-\nabla f(\mathbf{x}_k + \Delta \mathbf{x}_{k+1}) = H \Delta \mathbf{x}_{k+1} + \nabla f(\mathbf{x}_k)$$

Briefly describe the simplex method.

Introduce artificial variables. Simplices are not actually used in the method, but one interpretation of it is that it operates on simplicial cones, and these become proper simplices with an additional constraint. The simplicial cones in question are the corners (i.e., the neighborhoods of the vertices) of a geometric object called a polytope. The shape of this polytope is defined by the constraints applied to the objective function.

See lecture slides.

Briefly describe simulated annealing.

Introduce artificial temperature. Go to neighbouring state with probability 1 if the energy (or cost function) is lower there, and with probability $e^{-\Delta E/T}$ if the energy is higher.

Lower temperature with time.

Briefly describe evolutionary/genetic algorithms.

Encode information about the system in strings. Define genetic operations such as mutation, crossover and selection (various kinds exist). For every new time-step (generation), update the current population by mutating, and breeding new individuals. Then, finally, update the population according to how you have defined the selection rules for survival.

What is a cost function?

Something which determines the fitness of your system (how good it is). For example, how good is the fit to data given some well-defined measure?

Example: Least-squares distance.

Example: Cross-entropy.

How could a cost function be constructed? Can you think of some alternatives?

ASK CARL?

What is overfitting?

When your solution is too dependent on single data values. Example: prediction of outliers. In neural networks -- overtraining. Generally: non-generalizability.

What is sensitivity analysis, and why does it matter?

How sensitive is your system to small changes in your parameter (model) values?

1. Testing the robustness of the results of a model or system in the presence of uncertainty.
2. Increased understanding of the relationships between input and output variables in a system or model.
3. Uncertainty reduction: identifying model inputs that cause significant uncertainty in the output and should therefore be the focus of attention if the robustness is to be increased (perhaps by further research).
4. Searching for errors in the model (by encountering unexpected relationships between inputs and outputs).
5. Model simplification : fixing model inputs that have no effect on the output, or identifying and removing redundant parts of the model structure.
6. Enhancing communication from modelers to decision makers (e.g. by making recommendations more credible, understandable, compelling or persuasive).
7. Finding regions in the space of input factors for which the model output is either maximum or minimum or meets some optimum criterion (see optimization and Monte Carlo filtering).
8. In case of calibrating models with large number of parameters, a primary sensitivity test can ease the calibration stage by focusing on the sensitive parameters. Not knowing the sensitivity of parameters can result in time being uselessly spent on non-sensitive ones.

Why may some parameters be hard to determine?

Lack of knowledge about the range. May not be necessary for the system to function. Might be completely artificial and have no physical significance (interpretation hard). Also the counter-case -- the model might not be able to capture the complexity of the system.

How can you validate a model (and/or its parameters)?

Take out validation sample from data. Use cross-validation.

What makes a model good/useful?

It predicts.