Computational and Systems Biology Course 186—Modeling of Biological Systems by Connecting Biological Knowledge and Intuition with Mathematics and Computing

Instructor: Van Savage
Winter 2017 Quarter
Monday and Wednesday, 2-3:50pm
1/11/2017
Joke

When the apple fell and hit Newton on his head and he developed his law of gravitation, what if he’d been thinking about natural selection instead?
No office hours this week

Make up office hours on Tuesday of next week from 11am-12pm.
Fun Problem

- You have 9 balls. Eight are the same weight and one is heavier. You have a balance scale. How many weighings do you need to find the heaviest ball? What if you only knew one ball weighed differently but not more? How many weighings would it take?
How to model

- Draw picture and create notation
- Usually trying to describe change with time or space or some other dimension/variable
- Let’s start with Growth: basic biological process that shows up everywhere and allows us to review lots of math
- Discrete equations
- Calculus and Differential Equations
- Translating between two teaches a lot about implicit assumptions, potential numerical algorithms and computational approaches, and more
Growth and Branching—bacterial population
Growth and Branching—population abundance
Growth and Branching—vascular networks

Body size changes network size
Growth and Branching—human head and torso

Growth and Branching—mouse lung

Kristina Bostrom’s data
Growth and Branching—Tree
Growth and Branching—Fungi
Root of a gene tree captured in very few samples

Want to predict total number at each time point or particular paths going forward or backward?
Growth and Branching—Species/Taxonomy/Phylogeny
Equation for change: one discrete step

Translate idea/picture/process to equations

\[ x(t + 1) = 2x(t) \]

\[ x(t + 1) = r_d x(t) \]

\( r_d \) is an average that ignores individual variation.

\[ x(t + 1) = r_d x(t) + f(t) + \text{const} \]

Simple case in terms of biology and math but will see that several subtleties already arise.
How to solve this?
Steady state is easy to set up

\[ x = 2x \]

\[ x = r_d x \]

\[ x = r_d x + f(t) + \text{const} \]

But does not exist in any of these cases. Why? What would allow it?
Solve by iteration because this is just a recursion relation

\[ x(t + 1) = 2x(t) = 2 \cdot 2x(t - 1) = 2 \cdot 2 \cdot 2x(t - 2) = 2^{N+1} x(0) \]

Or just looking at process, one could guess this is solution

Recursion relations/rules are easy computationally. \( N \) is number of time steps or generations.

\[ x(t + 1) = r_d x(t) = r_d \cdot r_d x(t - 1) = r_d^{N+1} x(0) \]

Again, one could guess this by looking at previous charts or by plugging in \( k^t \) and solving for the constant \( k \).

Same if swapping space for time by sending \( t \) to \( x \).
What about for real units of time or space? How to convert this to differences and derivatives for arbitrary or small step sizes? Don’t implicitly assume step size must be 1.

\[
x(t + 1) - x(t) = (r_d - 1)x(t)
\]

\[
\lim_{\Delta t \to 0} \frac{x(t + \Delta t) - x(t)}{\Delta t} = \frac{(r_d - 1)}{\Delta t} x(t)
\]

\[
\frac{dx}{dt} = rx(t)
\]

\(r_d\) is number and \(r\) is rate. Keep units straight!
What about for real units of time or space? How to convert this to differences and derivatives?

\[ x(t) = x(0)e^{rt} \]

continuous solution

\[ x(N) = x(0)e^{\ln(r_d)N} = x(0)e^{\left[ \frac{\ln(r_d)}{\Delta t} \right] (N\Delta t)} \]

match to discrete solution?

Note we used \( N \), number of steps, not \( t \) for time (how we started) for discrete solution because that is what we really mean. Need to multiply and divide by \( \Delta t \) within exponential

Check for consistency at setup and solution. Doesn’t match. Is \( r \) related to \( r_d-1 \) or \( \ln(r_d) \)?
Instead of differences, why not think about ratios? Advantage is it gets rid of some unit choices

\[
\frac{x(t+1) - x(t)}{x(t)} = \frac{(r_d - 1)x(t)}{x(t)} = (r_d - 1)
\]

\[
\frac{1}{x(t)} \frac{dx}{dt} = \frac{d \ln x}{dt} = r
\]

Same equation as continuous one before.
What about taking logarithm of ratios at earlier stage? Becomes difference of log’s.

\[
\ln \left[ \frac{x(t+1)}{x(t)} \right] = \ln[x(t+1)] - \ln[x(t)] = \ln[r_d]
\]

\[
d\ln x = \frac{\ln(r_d)}{\Delta t}
\]

Now continuous eqn and solution are both same form as discrete solution. Why is this derivation more consistent?

Growth is exponential so log space is more natural space in which to think about change. That is linear “steps” and “differences” happen in ln space, not linear space. IMPLICIT CHOICE OF FUNCTIONAL SPACE IS ESSENTIAL AND REQUIRES THOUGHT! Can matter for computational choices and algorithms as well!
Exponential is linear in logarithmic space

Log transform changes exponential into linear space. Since solution is exponential, this means it changes our equation into linear space. Maybe hard to guess this first time so need **multiple iterations of thinking** to get it correct!
Also expressions all match in right asymptotic limit.

Taylor series and approximation

$$f(x) = \sum_{i=0}^{\infty} \left. \frac{d^i f(x)}{dx^i} \right|_{x=x^*} \frac{(x-x^*)^i}{i!}$$

What does each term tell you about shape? Why do we need infinitely many? How does distance from $x^*$ matter?
Get some match by expanding $\ln(1+x)$

$$\ln(1+x) \sim x$$

so

$$\ln(r_d) = \ln(1+(r_d-1)) \sim (r_d-1)$$

works when $x$ or $r_d-1$ is small because this is where $\ln$ steps look like linear steps and vice versa.
Also, what about small $\Delta t$ without limit all the way to 0?
What can we learn from this?

\[
x(t + \Delta t) = x(t) + \frac{dx}{dt} \bigg|_{\Delta t \ll t} + \frac{d^2x}{dt^2} \bigg|_{\Delta t \ll t} (t + \Delta t - t) + \frac{d^3x}{dt^3} \bigg|_{\Delta t \ll t} \frac{(t + \Delta t - t)^2}{2!} + \frac{d^4x}{dt^4} \bigg|_{\Delta t \ll t} \frac{(t + \Delta t - t)^3}{3!} + \ldots
\]

\[
= x(t) + \frac{dx}{dt} \bigg|_{\Delta t \ll t} + \frac{d^2x}{dt^2} \bigg|_{\Delta t \ll t} \Delta t + \frac{d^3x}{dt^3} \bigg|_{\Delta t \ll t} \frac{\Delta t^2}{2!} + \frac{d^4x}{dt^4} \bigg|_{\Delta t \ll t} \frac{\Delta t^3}{3!} + \ldots
\]

We were effectively just using first two terms on right side
for all derivations so far, which are only terms that exist in limit to
zero, but for any non-zero value of $\Delta t$ (all computational
methods), the second derivative term and higher orders matter
and choices about space and lack of linearity (i.e., curvature)
matter. We were implicitly ignoring higher-order terms for linear
space derivation before. For this case log space is where finite
steps in $\Delta t$ do not enter less because higher-order derivatives are
all zero because function is just linear. These exact
considerations are the core of numerical/computational
algorithms we will consider later.
Modeling process

Get idea as mental cartoon or process

Translate to equations

Solve equations

Check intuition and consistency of equations from start to finish

Consider if path to solution or intuition is easier in some other mathematical space.

Natural space for us hypothesizing about world is based on brain and evolution of how we process the world.

Main magic of mathematics is not just calculation but to translate to space that is natural for problem whether or not natural for our brain. Makes analytical, numerical, iterative, analysis, etc, much simpler

Repeating this process and gaining experience and exposure can change your intuition and natural space for thinking of brain. Math is language for doing this. Requires real thinking and not just application of recipes to use this language for thought and not parroting. This is the core of modeling.
Other basic pieces of translating biological idea and first discrete equations to calculus and differential equations and integration you’ve learned
What are higher-order derivatives in discrete form?

\[ D^2 x(t) = \frac{x(t + \Delta t) - x(t)}{\Delta t} - \frac{x(t) - x(t - \Delta t)}{\Delta t} \]

\[ = \frac{x(t + \Delta t) - 2x(t) + x(t - \Delta t)}{(\Delta t)^2} \]

Second derivative is change of change so above definition is natural. Higher-order derivatives are exact same idea, just apply change operation k times for kth derivative. Could have choice about two steps forward or two steps back and again this matters for numerical algorithms. Cannot subtract same step from self.

When doing any problem you should make friends with the problem, get to know it personally and not just apply a recipe, make all possible choices deliberately and in ways to bring out the best in your friend, and this will lead to much better and easier solutions! Don’t take your friend surfing if they can’t swim or give them chocolate if they like vanilla.
What about integration?

Already started thinking about this actually by talking about series. Riemann integration is one type of series approximation to integral

\[
\int f(x)\,dx = \sum_{i=0}^{N_{\text{steps}}} \frac{x_{\text{max}}}{\Delta a} \quad \text{Area}(i) = \lim_{\Delta a \to 0} \sum_{i=0}^{N_{\text{steps}}} f(i\Delta a) \cdot \Delta a
\]

\[ f(i\Delta a) \text{ maps to } f(x) \text{ and } \Delta a \text{ maps to } dx \]
Example 1: linear sum
(Gauss’ kindergarten problem)

\[ \sum_{i=0}^{N_{\text{steps}}} i = \frac{N(N+1)}{2} \]

Use this to do integral.

\[ \int x \, dx = \sum_{i=0}^{N_{\text{steps}}} (i \Delta a) \Delta a = (\Delta a)^2 \sum_{i=0}^{N_{\text{steps}}} i = (\Delta a)^2 \frac{N_{\text{steps}} (N_{\text{steps}} + 1)}{2} \]

\[ = (\Delta a)^2 \frac{\Delta a (\Delta a + 1)}{2} = \frac{1}{2} x_{\text{max}} (x_{\text{max}} + \Delta a) \sim x_{\text{max}}^2 \]

This is correct answer. Note that continuous/calculus form is simpler than discrete/sum form! Benefit of calculus is that it is simple! Equivalent to approximations that make math easy by ignoring lots of other terms that are extremely small or not measurable.
Example 2: Use partial fractions

\[ \sum_{i=1}^{N_{\text{steps}}} \frac{1}{i(i+1)} = \sum_{i=1}^{N_{\text{steps}}} \left( \frac{1}{i} - \frac{1}{i+1} \right) = \frac{N}{N+1} \]

Use this to do integral.

\[ \int_{0}^{x_{\text{max}}} \frac{1}{(1+x)^2} \sim \sum_{i=1}^{N_{\text{steps}}} \frac{1}{(1+i\Delta a)((1+i\Delta a)+\Delta a)} \Delta a = 1 - \frac{1}{(x_{\text{max}} + 1)} \]

Can more generally do any inverse power of x.
What about when inverse power is 1?

\[ \int_{0}^{x} \frac{1}{(1+x)} dx = \ln(1+x) = \sum_{i=1}^{N_{\text{steps}}} \frac{x^{i}}{i} \]

Latter is not obvious from above because trick is to take steps in log space or to log bin. Not linear space.
Ex. 3: Geometric sums and back to logarithms

$$\sum_{i=s}^{N} x^i = \frac{x^s - x^{N+1}}{1-x}$$

Use this to do integral.

$$\sum_{i=0}^{N_{\text{steps}}} x^{s+i\Delta a} \Delta a = \sum_{k'=s}^{N} x^{k'} \Delta k' \rightarrow \int_{s}^{N} x^k dk = \frac{x^N - x^s}{\ln(x)} \sim \frac{x^s - x^N}{1-x}$$

More generally,

$$\sum_{i=1}^{N_{\text{steps}}} f(x_{\text{min}} + i\Delta a) \Delta a = \sum_{k'=x_{\text{min}}}^{x_{\text{max}}} f(k') \Delta k' \rightarrow \int_{x_{\text{min}}}^{x_{\text{max}}} f(k)dk \rightarrow \sum_{k'=x_{\text{min}}}^{x_{\text{max}}} f(k')$$
Return to growth: cannot grow forever

Nature study in 1960 said human population growth is faster than exponential and called it super exponential. Said population would be infinite by 2026 if continued at the pace of growth in 2026.
Return to growth: cannot grow forever

\[
\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right)
\]

logistic growth

\[
\int_{N(0)}^{N(t)} \frac{K}{N(K-N)} dN = \int_{0}^{t} r \, dt \Rightarrow \int_{N(0)}^{N(t)} \left(\frac{1}{N} + \frac{1}{K-N}\right) dN = rt
\]

solving using partial fractions (deep part of logistic)

\[
N(t) = \frac{N(0)}{\left(1 - \frac{N(0)}{K}\right)e^{-rt} + \frac{N(0)}{K}}
\]

check limits at t=0 and t=\text{Infty}
Fixed points (no flow) now exist

\[
\frac{dN}{dt} = 0 = rN \left( 1 - \frac{N}{K} \right)
\]

\[N(t) = 0\]
\text{start}

\[N(t) = K\]
\text{finish}

At early times, solution reduces to same form as without \(K\) because do not see it yet

\[N(t) = N(0)e^{rt}\]
Plot of logistic growth

N(t)

K

t