No Atoms

So far
• atoms → coarse grained → lattices

Today – the holistic lecture
• from reaction kinetics to substitution matrices

What if we forget atoms and residues?
• Kinetics / dynamic systems
  • A → B breakdown of A, \( \frac{d[A]}{dt} = k[A] \)
  • foxes and hares \( \frac{dn_h}{dt} = \alpha n_h - \beta n_h n_f \) and \( \frac{dn_f}{dt} = \gamma n_h n_f - \delta n_f \)
  \( n_h, n_f \) number of hares and foxes
  • complicated kinetics – bacterium eats 10 different nutrients, makes 10 waste products, interconversion of nutrients

Andrew Torda July 2019, Struct and Sim
Plan

- simplest systems
  - one or two reactants
- treatment of more complicated systems
- transition matrices in sequences

Different approach next lecture
- handling very low probabilities
Simplest systems

- one species breakdown / radioactive decay
- A→B or A disappears
- philosophy
- we know the average disappearance of A

- Each molecule has an equal chance of breaking down: \( \frac{dA}{dt} = -kA \)

\[
\frac{dt}{dA} = -\frac{1}{kA}
\]

\[
t = -\frac{1}{k} \ln A - \frac{1}{k} \ln c
\]

\[
\ln A - \ln c = -kt = \ln \left( \frac{A}{c} \right)
\]

\[
\frac{A}{c} = e^{-kt} \quad \text{so} \quad A = ce^{-kt} \quad \text{or} \quad A = A_0e^{-kt} \quad \text{not unexpected}
\]
forward and backward reactions

\[
\begin{align*}
&k_1
\quad 2A \rightleftharpoons B \quad \text{so} \quad 2A \rightarrow B \quad \text{and rate of disappearance is} \quad k_1A^2, \quad \text{rate of appearance is} \quad k_2B \\
&k_2
\end{align*}
\]

\[
\begin{align*}
\frac{dA}{dt} &= -2k_1A^2 + 2k_2B \\
\text{and} \quad \frac{dB}{dt} &= k_1A^2 - k_2B
\end{align*}
\]

Theme

lots of processes are easiest to describe in differential form (rate of change)
These are easy enough to do by hand
Make it more complicated
An enzymatic reaction
\[ \begin{align*}
E + S & \rightleftharpoons ES \\
& \rightarrow E + P
\end{align*} \]

\[
\frac{dE}{dt} = -k_1 E \cdot S + k_2 ES + k_3 ES
\]

\[
\frac{dS}{dt} = -k_1 E \cdot S + k_2 ES
\]

\[
\frac{dES}{dt} = k_1 E \cdot S - k_2 ES - k_3 ES
\]

\[
\frac{dP}{dt} = k_3 ES
\]

let us rewrite..
\[
\frac{dE}{dt} = -k_1 E \cdot S + k_2 ES + k_3 ES \\
\frac{dS}{dt} = -k_1 E \cdot S + k_2 ES \\
\frac{dES}{dt} = k_1 E \cdot S - k_2 ES - k_3 ES \\
\frac{dP}{dt} = k_3 ES
\]

We have a matrix form
What is \( k_1 E \cdot S \) ? (and next terms)
you would usually say velocity vector \( \mathbf{v} \)
we can describe everything as \( \mathbf{s} = \mathbf{Nv} \)
General approach to kinetics

- "differential form" of kinetics
- applicable to most reactions

How is it helpful?

- $\frac{dA}{dt}$ is a velocity in one dimension
- velocity of A depends on where A is, B is, ...
- how to predict behaviour of system?

For some initial $A_t$ say $A_{t+\Delta t} = A_t + v\Delta t = A_t + \frac{dA}{dt}\Delta t$

- numerical integration exactly as in Newtonian dynamics
- do the same for $A, B, C$ ...

- Not just in this lecture – maple, matlab, deSolve in R, ..
Even more general

- We have a number of states $i, j, \ldots$ starting materials, products, intermediates
- We have a finite amount of material
  - Use the term probability $p_i$ for convenience and consistency
- $p_i(t + \delta t)$ depends on initial value, flux in and flux out

$$p_i(t + \delta t) = p_i(t) + \delta t \sum_{i \neq j} k_{ji} p_j(t) - \delta t \sum_{i \neq j} k_{ij} p_i(t)$$

$k_{ab}$ is rate constant for $a \rightarrow b$

- Or given a set of reactants and a matrix of $k$'s (rate matrix)
  - We can model the system
- If we say $v_{ij} = p_i k_{ij}$ what is the meaning of equilibrium? Every $v_{ij} = v_{ji}$
  - For an arbitrarily complicated system
    - I can find the set of $p$ ... equilibrium concentrations
the master equation

In chemical modelling, physical processes, work with master equation

Modelling in engineering
• put all components and possible routes into numerical bucket
• find steps which are bottle-necks
• effect of alternative pathways, think of multitude of protein folding pathways

Last property
• the state at \( t + \delta t \) depends on state at \( t \) and rate constants
• no dependence on previous states = Markov process

• what is the connection to sequences and mutations?
Markov processes and mutations

First – more general idea of transition matrices / Markov Chains

My system is described by a vector of probabilities – think amino acids at a site

\[
\mathbf{p} = \begin{bmatrix} p_A \\ p_G \\ p_C \\ \ldots \end{bmatrix}
\]

for ala, gly, cys, ...

\(p_{AB}\) probability of a transition AB but we have lots of them
A Markov transition matrix

\[
\begin{array}{cccc}
D & E & \cdots & W \\
D & p_{DD} & p_{DE} & \cdots & p_{DW} \\
E & p_{ED} & p_{EE} & \cdots & p_{EW} \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
W & p_{WD} & p_{WE} & \cdots & p_{WW}
\end{array}
\]

Only valid for short times
- \( D \rightarrow E \) OK
- \( D \rightarrow S \rightarrow T \rightarrow A \rightarrow D \rightarrow E \) something different

In Markov / probability framework rows sum to 1
Applying a matrix

- imagine three kinds of amino acid, $P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$
- population $E, D, W = 0.4, 0.4, 0.2$
- at time $t + \delta t$
  \[
  \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.4 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.7 \cdot 0.4 + 0.2 \cdot 0.4 + 0.1 \cdot 0.2 \\ 0.3 \cdot 0.4 + 0.6 \cdot 0.4 + 0.1 \cdot 0.2 \\ 0.1 \cdot 0.4 + 0.1 \cdot 0.4 + 0.8 \cdot 0.2 \end{bmatrix}
  \]
- gives us the new state of the system
- is this a substitution matrix?
comparison with a substitution matrix

blosum62:

|   | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
| A | 4 | -1 | -2 | -2 | 0 | -1 | -1 | 0 | -2 | -1 | -1 | -1 | -2 | -1 | 1 | 0 | -3 | -2 | 0 |
| R | -1 | 5 | 0 | -2 | -3 | 1 | 0 | -2 | 0 | -3 | -2 | 2 | -1 | -3 | -2 | -1 | -1 | -3 | -2 | -3 |
| N | -2 | 0 | 6 | 1 | -3 | 0 | 0 | 0 | 1 | -3 | -3 | 0 | -2 | -3 | -2 | 1 | 0 | -4 | -2 | -3 |
| D | -2 | -2 | 1 | 6 | -3 | 0 | 2 | -1 | -1 | -3 | -4 | 1 | -3 | -3 | -1 | 0 | -1 | -4 | -3 | -3 |
| C | 0 | -3 | -3 | -3 | 9 | -3 | -4 | -3 | -3 | -1 | 1 | -1 | -3 | -1 | -2 | -3 | -1 | -1 | -2 | -2 |
| Q | -1 | 1 | 0 | 0 | -3 | 5 | 2 | -2 | 0 | -3 | -2 | 1 | 0 | -3 | -1 | 0 | -1 | -2 | -1 | -2 |
| E | -1 | 0 | 0 | 2 | -4 | 2 | 5 | -2 | 0 | -3 | -3 | 1 | -2 | -3 | -1 | 0 | -1 | -3 | -2 | -2 |
| G | 0 | -2 | 0 | -1 | -3 | -2 | -2 | 6 | -2 | -4 | -4 | -2 | -3 | -3 | -2 | 0 | -2 | -2 | -3 | -3 |
| H | -2 | 0 | 1 | -1 | -3 | 0 | 0 | -2 | 8 | -3 | -3 | -1 | -2 | -1 | -2 | -1 | -2 | -2 | 2 | -3 |
| I | -1 | -3 | -3 | -3 | -1 | -3 | -3 | -4 | -3 | 4 | 2 | -3 | 1 | 0 | -3 | -2 | -1 | -3 | -1 | 3 |
| L | -1 | -2 | -3 | -4 | -1 | -2 | -3 | -4 | -3 | 2 | 4 | -2 | 2 | 0 | -3 | -2 | -1 | -2 | -1 | 1 |
| K | -1 | 2 | 0 | -1 | -3 | 1 | 1 | -2 | -1 | -3 | -2 | 5 | -1 | -3 | -1 | 0 | -1 | -3 | -2 | -2 |
| M | -1 | -1 | -2 | -3 | -1 | 0 | -2 | -3 | -2 | 1 | 2 | -1 | 5 | 0 | -2 | -1 | -1 | -1 | -1 | 1 |
| F | -2 | -3 | -3 | -3 | -2 | -3 | -3 | -3 | -1 | 0 | 0 | -3 | 0 | 6 | -4 | -2 | -2 | 1 | 3 | -1 |
| P | -1 | -2 | -2 | -1 | -3 | -1 | -1 | -2 | -2 | -3 | -3 | -1 | -2 | -4 | 7 | -1 | -1 | -4 | -3 | -2 |
| S | 1 | -1 | 1 | 0 | -1 | 0 | 0 | 0 | -1 | -2 | -2 | 0 | -1 | -2 | -1 | 4 | 1 | -3 | -2 | -2 |
| T | 0 | -1 | 0 | -1 | -1 | -1 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | -2 | -1 | 1 | 5 | -2 | -2 | 0 |
| W | -3 | -3 | -4 | -4 | -2 | -2 | -3 | -2 | -3 | -3 | -1 | 1 | -4 | -3 | -2 | 11 | 2 | -3 |
| Y | -2 | -2 | -2 | -3 | -2 | -1 | -2 | -3 | 2 | -1 | -1 | -2 | -1 | 3 | -3 | -2 | -2 | 2 | 7 | -1 |
| V | 0 | -3 | -3 | -3 | -1 | -2 | -2 | -3 | -3 | 3 | 1 | -2 | 1 | -1 | -2 | -2 | 0 | -3 | -1 | 4 |
where do blosum and PAM come from?

Take related sequences – no alignment errors

Count mutations (transitions) for each AB pair

Correct for the amount of A, B \((p_A, p_B, \ldots)\)
transition matrix versus blosum (PAM, JTT, Gonnet, ..)

Philosophically related – slightly different

• a substitution matrix is a log-odds creation - \( \log \frac{n^{obs}_{AB}}{n^{exp}_{AB}} \)
  • scaling does not matter

• a transition matrix is based on formal probabilities
  • if we have a composition vector \( \mathbf{v} \) elements sum to 1
  • after multiplication, still sum to 1

Similarities ...

• application to longer times
longer times

• transition matrix tells me about some change $\Delta t$
  
  $\mathbf{p}_{t+\delta t} = \mathbf{P}\mathbf{p}_t$ for composition vector $\mathbf{p}$ and matrix $\mathbf{P}$

• then at next time

  $\mathbf{p}_{t+2\delta t} = \mathbf{P}\mathbf{p}_{t+\delta t}$ or $\mathbf{P}\mathbf{P}\mathbf{p}_{t+\delta t}$

• to go to longer times, repeatedly multiply the matrix

• what happens? diagonal elements represent conservation ($p_{AA}$)
  
  • probability mass moves away from diagonal

• basis of PAM 100, PAM 200 ... substitution matrices

• when doing alignments, one should use the correct substitution matrix
infinite time

• I have a system described by probability of states \( \mathbf{p} \)
• I repeatedly multiply by a realistic \( \mathbf{P} \) ... \( \mathbf{P}^\infty \mathbf{p} \)

• does my distribution disappear? become flat?

• with infinite time everything becomes equally likely

• realistic? No
  • alignments become less reliable with evolutionary time
Summary so far

- chemical kinetics, mutation trajectories, fox + hare populations
  - examples of dynamic systems – very similar methods to treat them
  - allows one to treat complicated kinetics
  - usually simulated by numerical integration

- systems biology problems ? the same ?
  - sometimes  yes – sometimes neglect conservation of mass and formal treatment

- a Markov process state at $t + \delta t$ depends on state $t$
  - do not talk about second order or $n^{th}$ order processes

- everything so far depends on bulk properties
  - what happens if you only have a few molecules ? small numbers ? Last lecture
Systems with low numbers and noise

Usually we work with average properties

• averages require large numbers

When are they definitely wrong?

Does a system asymptotically approach equilibrium?
\frac{1}{6} \text{ chance of going backwards (away from equilibrium)}

Stock market
- yesterday trade at € 10
- buy offer at € 9, sell at € 11
- widow decides to sell husbands shares at € 9
- report of 10% share price drop
- you are asked to judge the significance
  - simulate how often it happens by chance

Queuing simulations
- shops, transport
more low copy dynamic systems

Lotka-Volterra

- foxes and hares \( \frac{dn_h}{dt} = \alpha n_h - \beta n_h n_f \) and \( \frac{dn_f}{dt} = \gamma n_h n_f - \delta n_f \)
- \( n_h, n_f \) number of hares and foxes
- but what if fox/hare meetings are not so common?

Dilute chemistry?

- lac repressor - < 40 copies per cell – well studied, classic DNA regulator
- what are chances of a protein repressor drifting through a cell and finding exactly the right piece of DNA?

Epidemiology

- states – healthy, sick, immune
Simulating rare events

Two aspects

• when do events occur?
• what to do?
Frequencies of rare events

Events are not correlated
- this particle is independent of that one
- calls into help line are independent
- flood in this time not correlated with some other time

Average $\mu$ is known – number of events in a time period in time $t$
- average number of calls in day, Geiger counter counts / s, ..

- later use rate $\lambda$ so in time $t$, $\mu = \lambda t$
- average time between events? $\tau = \lambda^{-1}$ (check dimensions here)

Two names will keep coming up
- poisson distribution - think of $\mu$
- exponential distribution - think of $\tau$
Poisson

Used for next step

What is the probability of \( n \) events in time \( t \)?

\[
P(x = n) = e^{-\mu} \cdot \frac{\mu^n}{n!}
\]

\( \mu = 2 \) but probability of seeing 2 events is only \( \approx \frac{1}{4} \)

how to derive? Do derivation of binomial and take limit
time between events

We have $P(x = n) = e^{-\mu} \frac{\mu^n}{n!} = e^{-\lambda t} \frac{\mu^n}{n!}$

• something does not happen for $\tau$, then happens

• zero events over some $t$? $P(x = 0) = e^{-\lambda t} \frac{\mu^0}{0!} = e^{-\lambda t}$

This means the first event happened at $\tau$ and $\tau > t$ so

• $P(\tau > t) = P(x = 0) = e^{-\lambda t}$ but then probability of an event is

• $P(\tau \leq t) = 1 - P(x = 0) = 1 - e^{-\lambda t}$

• Cumulative probability over all $\tau$ is $1 - e^{-\lambda t}$

• instantaneous probability for some $t$ will be the derivative
time between events

Cumulative probability over all $\tau$ is $1 - e^{-\lambda t}$

Instantaneous probability for some $t$ will be the derivative

$$\frac{d}{dt} P(T \leq t) = \frac{d}{dt} (1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

- distribution of gaps between events $\tau$ is $\lambda e^{-\lambda t}$ or exponential distribution

Formally, $\tau$ is a random variable drawn from $f(\tau, t) = \lambda e^{-\lambda t}$

- back to simulation question
simulating with rare events

- $\lambda$ is 10 events a second or 20 calls an hour or..
- define our time step as $\tau$
  
  because $\tau$ is the time between events

Simulate

while ($t < t_{max}$)
  pick $\tau$ from $P(\tau = t) = \lambda e^{-\lambda t}$
  $t := t + \tau$
  do something

Bit more complicated
more than one event type

\[
\begin{align*}
  k_1 & \quad k_3 \\
  E+S & \rightleftharpoons ES \rightarrow E + P \\
  k_2 &
\end{align*}
\]

- three reactions – each is a poisson process
- total poisson process
  - I have A's and B's happening independently
  - I see \( \mu_A \) events and \( \mu_B \) events
    - total \( \mu_0 \) is just \( \mu_A + \mu_B \) so I can just add up \( \lambda \)'s

\[ P(x = n) = e^{-\mu} \frac{\mu^n}{n!} \] \( \mu \) is the average number of times something happens

- add up the rates, say \( \lambda_0 = \sum_{i=1}^{N\text{rates}} \lambda_i \)

\[ P(x = n) = e^{\lambda_0 t} \frac{\mu^n}{n!} \] or maybe you prefer \( P(x = n) = e^{-\mu_0} \frac{\mu^n}{n!} \)

- we can draw timestep from this distribution, but what happens there?
\[ \mu_0 = \lambda_0 t \quad \text{total events per time} \]

- simulate

while \((t < t_{max})\)
  pick \(\Delta t\) from \(P(\tau = t) = \lambda_0 e^{-\lambda_0 t}\)
  \(t := t + \tau\)
  pick a reaction

- choosing a reaction...
choosing a reaction

\( N_\lambda \) rates each \( \lambda_i \) (three in previous example)

- probability of reaction \( i \)

\[
p_i = \frac{\lambda_i}{\sum_{j=1}^{N_\lambda} \lambda_j}
\]

implementation to choose which reaction happens

- make a table of \( q_i = \sum_{j=1}^{i} p_j \)

\[
r = \text{rand}(0..1)
\]

\[
\text{for } (i = 0; i < n; i++) {
    \text{if } r < q_i \{ \text{return } i \}
}
\]

\[
\begin{array}{cc}
\hline
p_i & q_i \\
0.2 & 0.2 \\
0.3 & 0.5 \\
0.5 & 1.0 \\
\hline
\end{array}
\]
The Gillespie algorithm

• $\mu_0 = \lambda_0 t$  total events per time
• calculate rate $\lambda_0$

• simulate
while ($t < t_{\text{max}}$)
    pick $\tau$ from $P(\tau = t) = \lambda_0 e^{-\lambda_0 t}$
    $t: = t + \tau$
    pick a reaction from recipe on previous slide
    update rates ($\lambda$'s) since quantities have changed

What did Mr Gillespie find?
\[ \overline{X} + Y \overset{c_1}{\rightarrow} 2Y \] and \( \overline{X} \) means a large pool of X

\[ 2Y \overset{c_2}{\rightarrow} Z \]

Two starting conditions

- \( n_Y = 10 \) or 1000
  - some values for \( c_1, c_2 \) and \( n_X \)

Have you learnt much?

Why do Gillespie simulations?

You already know average behaviour from classic kinetics

- You can predict $[Y]_t$ but it is an average

Run simulation 1000 times

- gives you $[Y]_t \pm \sigma_Y$
- can predict fluctuation around equilibrium values
Gillespie-style Methods

Back to cell with one DNA + 40 copies of repressor
• from some estimates of kinetics, can predict
  • average occupancy
  • lifetime of bound state
  • fraction of time DNA site is occupied, confidence intervals

Stock exchange example
• you hear of a 10 % drop in share price – has something really happened?

These methods give you
• errors / fluctuations / significance / confidence intervals
Extensions / Applications of Gillespie method

Fuchsen + blue hares and red hares
- move randomly, meet randomly fox + hare $\rightarrow$ fox
  - widely used in eco-system modelling

Spatial diffusion problem in cells
- for a particle in box$_1$, box$_1 \mathbin{\xrightarrow{k}}$ box$_2$
  - diffusive simulations + chemistry – states are mixture of chemistry and location

Finance
- few sellers and buyers
Alternative philosophy

Follow a trajectory in some field

A particle moves is hit by other particles
  • you do not want to model the particles explicitly
  • a chemical trajectory with side reactions

but path is noisy
Adding noise to systems

Examples here

- Gaussian (normal) noise
  - mean $\mu = 0$
  - call my noise $W(t)$ means $\mu = 0$ and variance $(\sigma^2) = t$
  - not obvious – Brownian processes – you move randomly

Want to build noise into normal simulations

- Normally (Newtonian dynamics, chemical kinetics) – simple integrator
  \[
  \frac{dx}{dt} = f(x) \quad \text{where } f \text{ comes from a force or chemistry rate of change}
  \]
  and we have just said $x_{n+1} = x_n + \Delta t \ f(x)$

  use $W(x)$ – Wiener process
\[ x_{n+1} = x_n + \Delta t \, f(x) \] can also write
\[ x_{n+1} = x_n + dx \]

if I have a random process \( W \)
\[ x_{n+1} = x_n + dW \]

Meaning of \( dW \)?
• \( W(t) \) is the fluctuation over \( t \) – random variable from Gaussian \((0,t)\)
• \( dW \) also a random variable - \( \sqrt{\Delta t} \cdot \text{gaussian}(0, t) \) more concisely \( \sqrt{t}N(0, t) \)
  usually use \( N() \) to represent Gaussian random number
integrate over random variable

\[ \frac{dx}{dt} = f(x) \quad \text{so} \quad dx = f(x(t))dt \]

\[ x = \int f(x(t))dt \]

to

for random variable

\[ dx = dW \]

\[ x = \int dW \]

define \( X(T) = \int_0^T x(t)dW(t) \) make it discrete

\[ \lim_{\Delta t \to 0} \sum_{j=0}^{N_{step}} x(j\Delta t) \cdot (W((j+1)\Delta t) - W(j\Delta t)) = \sum_{j=0}^{N_{step}} x(j\Delta t) \left( \sqrt{\Delta t} N(0,1) \right) \]

- so a recipe for the diffusive / Brownian motion
- more interesting to combine it
A variable $X$ feels a deterministic force $f(X)$ and random $g(X)$

$$\frac{dX(t)}{dt} = f(X(t))dt + g(X(t))dW(t)$$

$$X(t) = X_0 + \int_0^t f(X(x))ds + \int_0^t g(X(s))dW(s)$$

- think of a protein *in vacuo* with Newtonian dynamics from $f(X)$ and random effects of solvent from $g(X)$

- connect back to last week and this week chemistry
stochastic chemistry – not Gillespie

\[ A + B \stackrel{\kappa}{\rightarrow} C \]
\[ dA = -kAB \ dt \quad dB = -kAB \ dt \quad dC = kAB \ dt \]

• then add noise
\[ dA = -kAB \ dt + \kappa A \ dW_1(t) \quad dB = -kAB \ dt + \kappa B \ dW_2(t) \quad \text{and} \]
\[ dC = kAB \ dt + \kappa C \ dW_3(t) \]

• simulating ? easy
\[ A_{i+1} = A_i - kA_i B_i \Delta t + \kappa A_0 \sqrt{\Delta t} \ N(0,1) \]
and similar for \( B \) and \( C \)
\[ A + B \rightarrow C \]

\[ k = 1 \quad \kappa = 0.01 \]

\[ A(0) = 1, \quad B(0) = 1, \quad C(0) = 0 \]

Who uses this?

- Chemistry – reactions with random side reactions
- Epidemiology
- Ecosystems
- Finance – first Google hit with maple .. finance [wienerprocess]
Ende Des Semesters

Last week
• simulations and processes using just a transition matrix
• from chemistry to mutations (also works for epidemiology, finance)

Gillespie
• very rigorous
• rather slow

Stochastic differential methods
• general noise
• Brownian dynamics, markets, epidemiology, chemical kinetics
• requires a model for noise - occasionally rigorous