Last time: serial dilution model of evolution

$N_1(t) = N_1(0)e^{rt}$

$N_2(t) = N_2(0)e^{(r+s)t}$

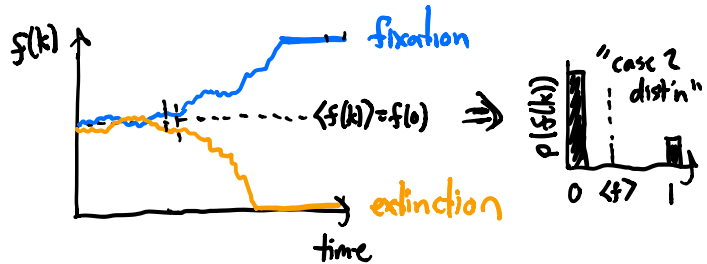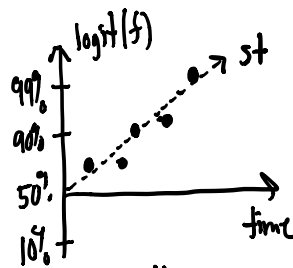$\underbrace{\phantom{N_2(t) = N_2(0)e^{(r+s)t}}}$
simple phenotype
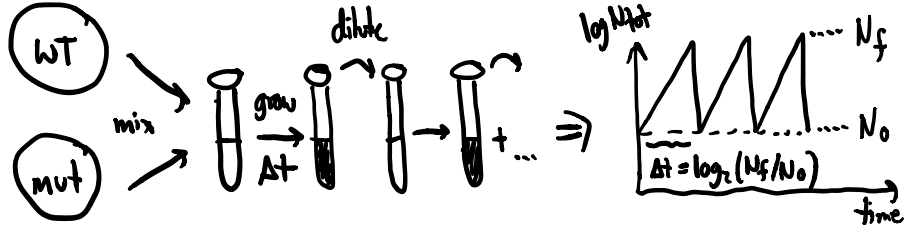
"Markov model" for frequency trajectory $(f(0), f(1), f(2), \dots)$

$f(k+1) = \dfrac{N_2}{N_2 + N_1}$

$N_2 \sim \text{Poisson}\left(N_0 \dfrac{f(k)e^{s\Delta t}}{f(k)e^{s\Delta t} + 1 - f(k)}\right)$

$N_1 \sim \text{Poisson}\left(N_0 \dfrac{1 - f(k)}{f(k)e^{s\Delta t} + 1 - f(k)}\right)$

---

① when $s = 0$:

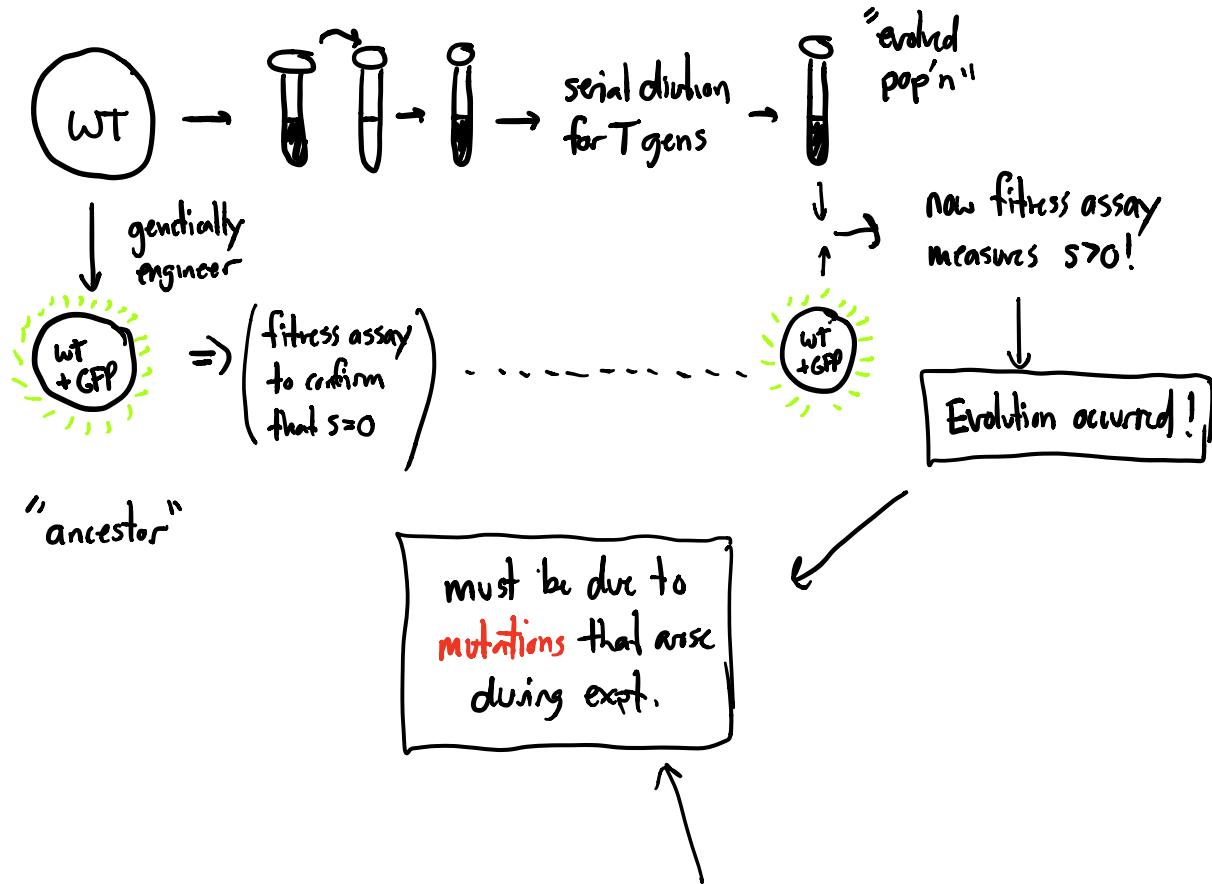$f(k+1) = f(k) \pm \Theta\left(\dfrac{1}{\sqrt{N_0}}\right)$

② when $s > 0$, $N_0 = \infty$:

$\log\left(\dfrac{f(t)}{1 - f(t)}\right) = \log\left(\dfrac{f(0)}{1 - f(0)}\right) + st$

Can measure s from measurements of $f(k)$ in short-term mixing expts.

"fitness assay"

Today: consider the following experiment:

WT → serial dilution for T gens → "evolved pop'n"

genetically engineer ↓

WT + GFP ⟹ (fitness assay to confirm that s=0)

"ancestor"

now fitness assay measures s>0!

WT* + GFP

Evolution occurred!

must be due to **mutations** that arose during expt.

Today: ① How to model this process
② "Microscopic models" & th "diffusion limit"

Start w/ simplest case: suppose there is a single ~~target~~ for mutations

(e.g. WT → ΔsugarX) that happens w/

probability μ per division $(\mu \ll 1)$

⟹ | this is called a "single locus" model |

equivalent to a genome w/ a single site

$$L=1$$

(will learn how to generalize to bigger genomes later...)

⟹ @ some time, start w/ 0 mutants in pop'n @ beginning of day:
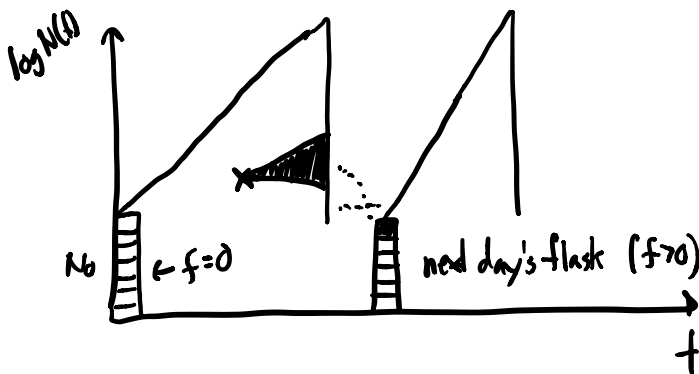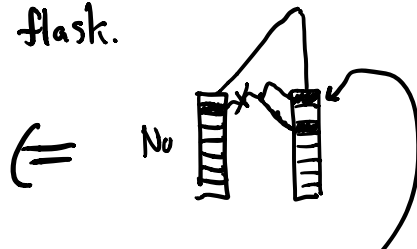

N_0 ← f=0   next day's flask (f>0) ⟹ then our existing model applies

For simplicity, assume:

① Fitness differences don't start until next day.

② every cell @ beginning of next day's flask traces back to cell alive @ beginning of previous day's flask.

by definition, $\Delta t = \log_2\left(\frac{N_f}{N_0}\right)$ divisions separate them.

$\Longleftarrow$ N_0

$\Rightarrow$ prob[mutation] $= \nu \cdot \Delta t$

<u>probability that this cell has acquired mutation.</u>

$\Rightarrow$ approx is that mutations occur in cells <u>independently</u>

$$N_2 \sim \text{Poisson}\left(N_0 \, p_{mut}\right)$$

$$N_1 \sim \text{Poisson}\left(N_0 (1 - p_{mut})\right)$$

"everyone who didn't mutate"

$\Rightarrow$ $f(k+1) = \dfrac{N_2}{N_2 + N_1}$ $\quad\left(\begin{array}{l}\text{then} \\ \text{previous} \\ \text{dynamics} \\ \text{apply}\end{array}\right)$

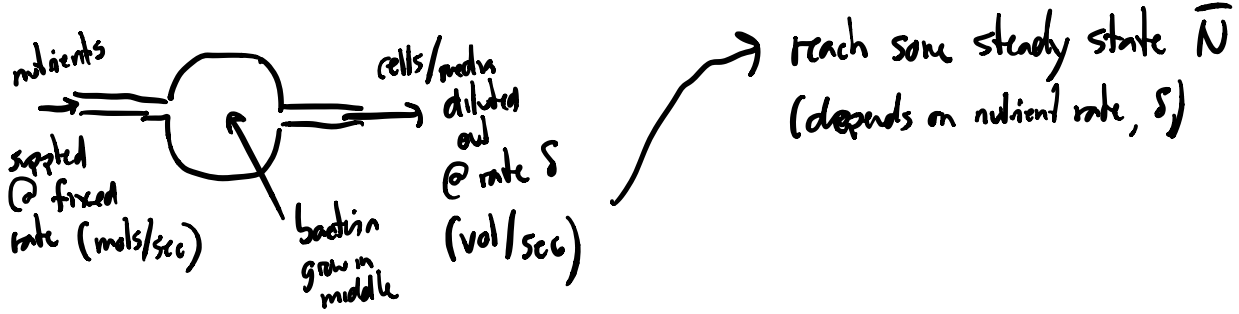$\Rightarrow$ "full model" ("microscopic model of serial dilution")

$$N_2 \sim \text{Poisson}\left(N_0 \, \frac{f(k) e^{s\delta t}}{f(k) e^{s\delta t} + 1 - f(k)}\right) + \text{Poisson}\left(N_0 \, p_{mut} \times \left(\frac{1 - f(k)}{f(k) e^{s\delta t} + 1 - f(k)}\right)\right)$$

$$N_1 \sim \text{Poisson}\left(N_0 (1 - p_{mut})\left(\frac{1 - f(k)}{f(k) e^{s\delta t} + 1 - f(k)}\right)\right)$$

$\left(\begin{array}{l}\text{can also add} \\ \text{"back mutations"} \\ \text{@ rate } \nu\end{array}\right)$

$\Rightarrow$ $f(k+1) = \dfrac{N_2}{N_2 + N_1}$
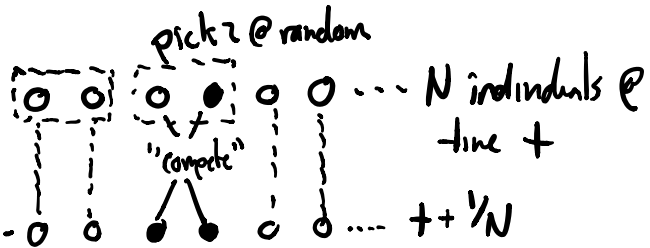
Microscopic models :   lots of them !

e.g. one experimentally motivated one ⇒ "a chemostat"

nutrients
→
supplied
@ fixed
rate (mols/sec)

bacteria
grow in
middle

cells/medium
diluted
out
@ rate S
(vol/sec)

→ reach some steady state $\bar{N}$
(depends on nutrient rate, S)

⇒ other "pure math" ones (population genetics)

"ball & urn models"   "bean bag genetics"

⇒ e.g. "Moran model"

pick 2 @ random

○ ○ ┊ ○ ● ┊ ○ ○ ··· N individuals @
                              time t
"compete"
● ● ··· $+ + 1/N$

$\Rightarrow$ one divides & replaces other
w/ prob $\dfrac{e^S}{1+e^S}$ → "fitness"

need N steps before
everyone has "reproduced" once

$\Rightarrow$ e.g. "Wright-Fisher model"

N individuals @ time t

time $t+1$
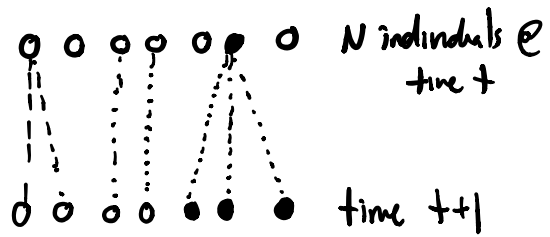
parent of each new guy is selected @ random w/ weight $\propto e^{S_i}$

when $s=0$
exact relation for:

$$E\left[f(t)\right] = E\left[f(t-1)\right] = f(0)$$

$$E\left[f(t)(1-f(t))\right] = \left(1-\frac{1}{N}\right)E\left[f(t-1)(1-f(t-1))\right] = f(0)(1-f(0))\,e^{-t/N}$$

$\underbrace{}$
"heterozygosity"
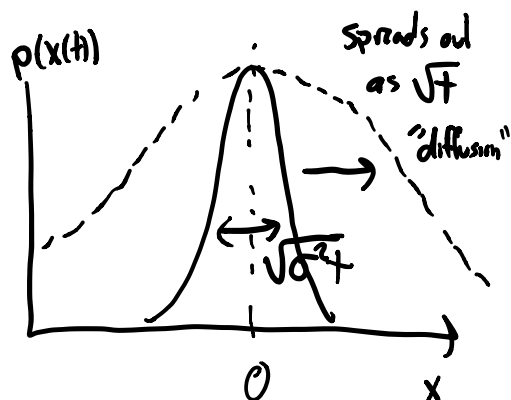
Simple math problem: discrete random walk.

Let $\Delta X_1, \Delta X_2, \ldots \sim \text{Gaussian}(0, \sigma^2)$

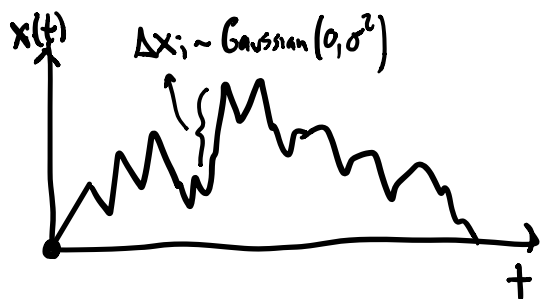& define: $X(t) = \sum\limits_{i=1}^{t} \Delta X_i \implies$ stochastic process from $X(t) \to X(t+1)$

(or in recursive form: $X(t+1) = X(t) + \Delta X_t$)

Recall: $\text{Gaussian}\left(\mu_1, \sigma_1^2\right) + \text{Gaussian}\left(\mu_2, \sigma_2^2\right)$

$$= \text{Gaussian}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

$$\Rightarrow \quad X(t) \sim \text{Gaussian}\left(0, \sigma^2 t\right)$$

$p(x(t))$

spreads out as $\sqrt{t}$

"diffusion"

$\sqrt{\sigma^2 t}$

0

x

can also consider an arbitrary path:

$x(t)$

$\Delta x_i \sim \text{Gaussian}\left(0, \sigma^2\right)$

t

$$p\left(x(0), x(1), \dots, x(t)\right) =$$

$$\prod_{i=1}^{t} \frac{e^{-\frac{(x(t) - x(t-1))^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

what if Δx_i not Gaussian?

$$\Rightarrow \text{now let } \Delta x_1, \Delta x_2, \dots \overset{iid}{\sim} p(\Delta x) \text{ w/ } \langle \Delta x \rangle = \mu$$
$$\text{Var}(\Delta x) = \sigma^2$$

Now:

$$X(t) = \Delta x_1 + \Delta x_2 + \Delta x_3 + \dots + \Delta x_{t-2} + \Delta x_{t-1} + \Delta x_t$$

$t \to \infty$ often, (by central limit theorem)

$\underbrace{}$

can also apply CLT
for sub-intervals of
length $\delta t \gg 1$ (at least)

Gaussian$(\mu t, \sigma^2 t)$

$\Longrightarrow$ $\delta x \sim$ Gaussian$\left(\mu \delta t, \sigma^2 \delta t\right)$

$\Rightarrow$ so if "coarse-grain" over timescale $\delta t \gg 1$

can rewrite our sum as

$$X(t) = \delta x_1 + \delta x_2 + \dots \delta x_{(t/\delta t)}$$

$\hookrightarrow \Delta x_1 + \Delta x_2 + \dots \Delta x_{\delta t} \sim$ Gaussian$\left(\mu \delta t, \sigma^2 \delta t\right)$

or, in recursive notation:

$$X(t + \delta t) = X(t) + \delta x(t) = X(t) + \mu \delta t + \sqrt{\sigma^2 \delta t}\, Z_t$$

independent
standard gaussian
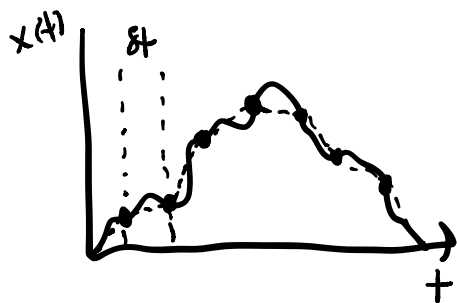$\Uparrow$ w/ $\langle Z_t \rangle = 0$
   Var$(Z_t) = 1$

or as stochastic differential equation (SDE)

$$\frac{dx}{dt} = \underbrace{\mu}_{\substack{\text{"deterministic"}\\ \text{part}}} + \underbrace{\sqrt{\sigma^2}\, \eta(t)}_{\text{"stochastic part"}}$$

$\to$ "Brownian noise term"

code for
this
series
expansion
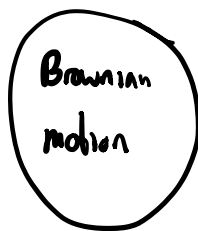
upshot: can write down probability of arbitrary path:

$$p\left(x(0), x(\delta t), x(2\delta t), \ldots\right)$$

$$= \prod_{i=1}^{(t/\delta t - 1)} \frac{1}{\sqrt{2\pi\sigma^2 \delta t}} \, e^{-\frac{\left(x((i+1)\delta t) - x(i\delta t)\right)^2}{2\sigma^2 \delta t}}$$

microscopic models:

$$P_1(\Delta x)$$

$$P_2(\Delta x)$$

$$\vdots$$

$$\text{Gaussian}(\mu, \sigma^2)$$

$\delta t \to 1$

Brownian motion

"universality class"

$$= \text{microscopic} + \text{limit} \ (\delta t \to 1)$$

or

("physical diffusion")