

Announcements: Problem Set 2 posted (DUE: 2/9/21)

Last time: Random walks & diffusion

$$\xrightarrow{\text{iid}} \rho(\Delta x) \text{ w/ } \langle \Delta x \rangle = \mu, \text{Var}(\Delta x) = \sigma^2$$

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

↓ "coarse-grain" over intermediate timescale $\delta t \gg 1$

$$\delta x_1 + \dots + \delta x_t \rightarrow \text{Gaussian}(\mu t, \sigma^2 t)$$

$$\xrightarrow{\quad} \text{Gaussian}(\mu \delta t, \sigma^2 \delta t)$$

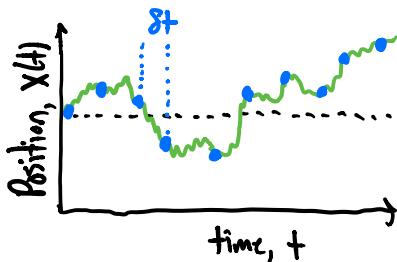
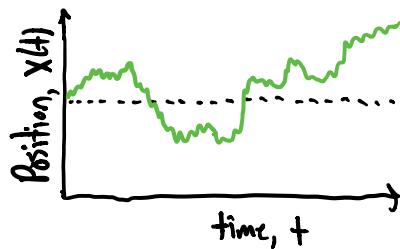
↓ will write as

$$X(t+\delta t) \sim X(t) + \mu \delta t + \sqrt{\sigma^2 \delta t} Z_t \sim N(0, 1)$$

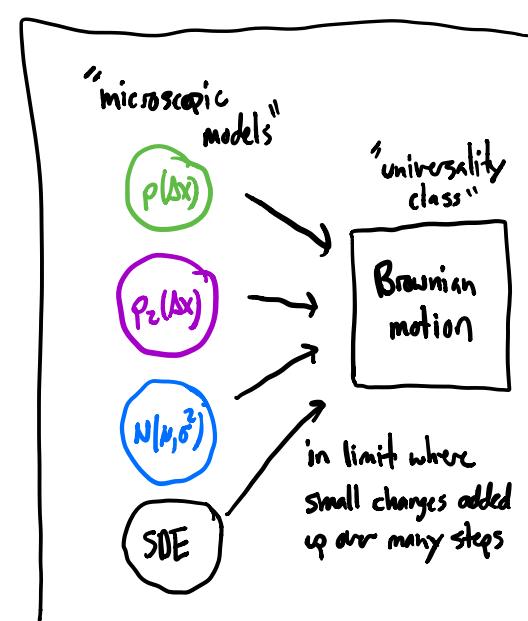
or

$$\text{SDE: } \frac{dx}{dt} = \mu + \sqrt{\sigma^2} \eta(t)$$

in this class,
shorthand for

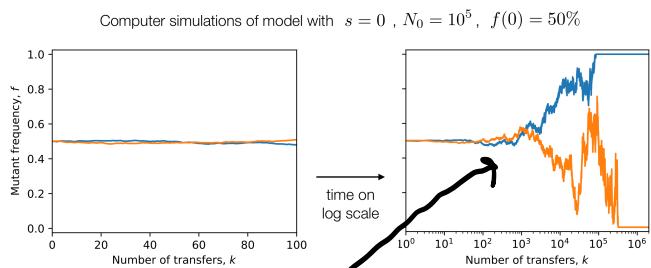
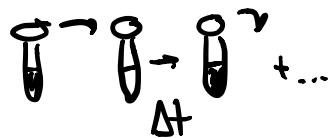


Today: ① Can this apply to evolution?
② Working w/ SDEs.



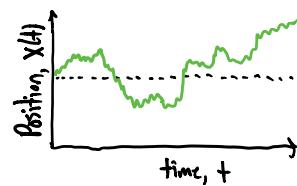
Basic idea: evolutionary phenomena take place over many generations...
 \Rightarrow maybe similar diffusion-like behavior emerges when $1 \ll s\tau \ll t$?

E.g. serial dilution model:



mutation frequency looks a bit like random walk...

\Rightarrow can we use same approach here?



$$f(t + \Delta t) = f_0 + \underbrace{\Delta f_1}_{\text{ }} + \underbrace{\Delta f_2}_{\text{ }} + \underbrace{\Delta f_3}_{\text{ }} + \dots + \underbrace{\Delta f_{k-2}}_{\text{ }} + \underbrace{\Delta f_{k-1}}_{\text{ }} + \underbrace{\Delta f_k}_{\text{ }} \quad (k = \frac{t}{\Delta t})$$

↓ ↓ ↓

coarse grain over
timescale $s\tau \gg \Delta t$

$$= f_0 + \delta f_1 + \dots + \delta f_{(\frac{t}{s\tau})}$$

in this case:

$$\underline{\Delta f(t)} \equiv f(t + \Delta t) - f(t) \equiv \frac{N_2}{N_2 + N_1} - f(t)$$

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

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

$$\delta f_1 = \Delta f_1 + \Delta f_2 + \dots + \Delta f_{\frac{\delta t}{\Delta t}} \rightarrow \text{can make this } \cancel{\xrightarrow{\quad}} \begin{array}{l} \text{can't apply} \\ \text{CLT} \\ \text{from before} \end{array}$$

$\mu(f(t))$
 $\sigma^2(f(t))$

$\mu(f(t+\Delta t)) \approx \mu(f(t))$
 $\sigma^2(f(t+\Delta t)) \approx \sigma^2(f(t))$

Key idea: if we coarse grain over many gens (for CLT) $\left[\begin{array}{c} \text{lower} \\ \text{bound} \\ \hline \Delta t \end{array} \right]$

* but sufficiently few gens s.t. $f(t+\Delta t) \approx f(t)$ $\left[\begin{array}{c} \text{upper} \\ \text{bound} \end{array} \right]$

\swarrow then $\mu(f(t+i\Delta t)) \approx \mu(f(t))$ for all $i \leq \frac{\delta t}{\Delta t}$
 $\sigma^2 \dots$

\Rightarrow can therefore apply CLT from before: $\left[\begin{array}{c} \text{t: } \left(\frac{\delta t}{\Delta t} \right) \end{array} \right]$

$$\begin{aligned} \delta f &= \Delta f_1 + \Delta f_2 + \dots + \Delta f_{\frac{\delta t}{\Delta t}} = \text{Gaussian} \left(\mu(f(t)), \downarrow, \sigma^2(f(t)) \delta t \right) \\ &= \mu(f(t)) \delta t + \sqrt{\sigma^2(f(t)) \delta t} Z_t \sim N(0, 1) \end{aligned}$$

\Rightarrow can we show that this works? (\Rightarrow when it works?)

\Rightarrow use self-consistency + series approximations. *

HW
Problem

Step 1: what limits are relevant? (e.g. $N_0, s, \text{etc.}$)

\Rightarrow focus on a single timestep:

(a) need $\text{Poisson}\left(\frac{N_0 f e^{s\Delta t}}{f e^{s\Delta t} + 1 - f}\right) \approx \text{Poisson}(N_0 f)$

\Rightarrow will be true when $s\Delta t \ll 1$

(b) need $\text{Poisson}(N_0 f) \approx N_0 f$

$\xrightarrow{\text{large } N_0} N_0 f \pm \sqrt{N_0 f} \Rightarrow f(t+\Delta t) = f \pm \sqrt{\frac{c}{N_0}}$

\Rightarrow will be true when $N_0 \gg 1$

(strictly speaking, $N_0 f \gg 1, N_0 f(1-f) \ll 1$)
 \hookrightarrow will discuss later.

Step 2: calculate leading order contributions to

$$\mu(f(t)) \equiv \langle \Delta f \rangle \text{ and } \sigma^2(f(t)) \equiv \text{Var}(\Delta f)$$

in single timestep (using limits above)

$$\begin{aligned} \textcircled{1} \text{ argument of Poisson: } & \frac{f(t)e^{s\Delta t}}{f(t)e^{s\Delta t} + 1-f(t)} \approx \frac{f(t)[1+s\Delta t+\dots]}{f(t)[1+s\Delta t+\dots] + 1-f} \\ & \approx f(1+s\Delta t)(1-fs\Delta t) \\ & \approx f + s\Delta t f(1-f) + \underbrace{h.o.t.}_{\text{leading order contribution}}. \end{aligned}$$

\textcircled{2} Gaussian approx to Poisson dist'n:

$$\begin{aligned} f(t+\Delta t) &= \frac{N_0[f+s\Delta t+f(1-f)] + \sqrt{N_0 f} Z_1}{(\dots) + N_0[1-f-s\Delta t+f(1-f)] + \sqrt{N_0(1-f)} Z_2} \\ &= \frac{f + s\Delta t f(1-f) + \sqrt{\frac{f}{N_0}} Z_1}{1 + \sqrt{\frac{f}{N_0}} Z_1 + \sqrt{\frac{1-f}{N_0}} Z_2} \end{aligned}$$

$$\begin{aligned}
 &\approx f + s\Delta t + f(1-f) + \underbrace{\sqrt{\frac{f}{N_0}} Z_1 - f\sqrt{\frac{f}{N_0}} Z_1 - f\sqrt{\frac{f-f}{N_0}} Z_2}_{\sqrt{\frac{(1-f)^2 f}{N_0}} Z_1 - \sqrt{\frac{f^2 (1-f)}{N_0}} Z_2} + \text{h.o.t.} \\
 &\quad \underbrace{\sqrt{\frac{(1-f)^2 f}{N_0}} Z_1 - \sqrt{\frac{f^2 (1-f)}{N_0}} Z_2}_{\sqrt{\frac{(1-f)^2 f}{N_0} + \frac{f^2 (1-f)}{N_0}}} Z_3 \\
 &f(t+\Delta t) \approx f + s\Delta t + f(1-f) + \underbrace{\sqrt{\frac{f(1-f)}{N_0}} Z}_{\mu(f)} \quad \left. \begin{array}{l} (\Delta t \ll 1) \\ (N_0 \gg \text{big}) \end{array} \right\} \\
 &\quad \downarrow \qquad \qquad \qquad \sigma^2(f) \quad (\text{that's one cycle})
 \end{aligned}$$

Step 3: add up contributions over Δt gens ($\frac{\Delta t}{\Delta t}$ cycles)

$$\delta f = s\Delta t + f(1-f)\left[\frac{\Delta t}{\Delta t}\right] + \sqrt{\frac{f(1-f)}{N_0}\left(\frac{\Delta t}{\Delta t}\right)} Z_+ \quad \left. \begin{array}{l} (\text{assumes CLT}) \\ (+ \text{homogeneity}) \end{array} \right\}$$

$$\begin{aligned}
 &= \underbrace{s f(1-f) \Delta t}_{\downarrow} + \sqrt{\frac{f(1-f) \Delta t}{(N_0 \Delta t)}} Z_+ \\
 &\quad \text{"s_e" effective selection strength} \qquad \qquad \qquad \text{"N_e" effective population size.}
 \end{aligned}$$

Step 4: check self consistency (important!!!)

① need $f(t+\delta t) \approx f(t) \Rightarrow \delta f \ll f \wedge \delta t \ll 1-f$

\Rightarrow a Need $s\delta t \ll 1 \Rightarrow \delta t \ll \frac{1}{s}$ (selection timescale)

b $\frac{\delta t}{N_0 \Delta t} \ll 1 \Rightarrow \delta t \ll N_0 \Delta t$ (drift timescale)

\Rightarrow ensures homogeneity w/in coarse grain step.

② $\delta t \gtrsim \Delta t$ from CLT condition.

($\delta t \gg \Delta t$)

\Rightarrow can satisfy both when



$s \rightarrow 0$
 $N_0 \rightarrow \infty$

while $N_0 s \Delta t$ can be anything.

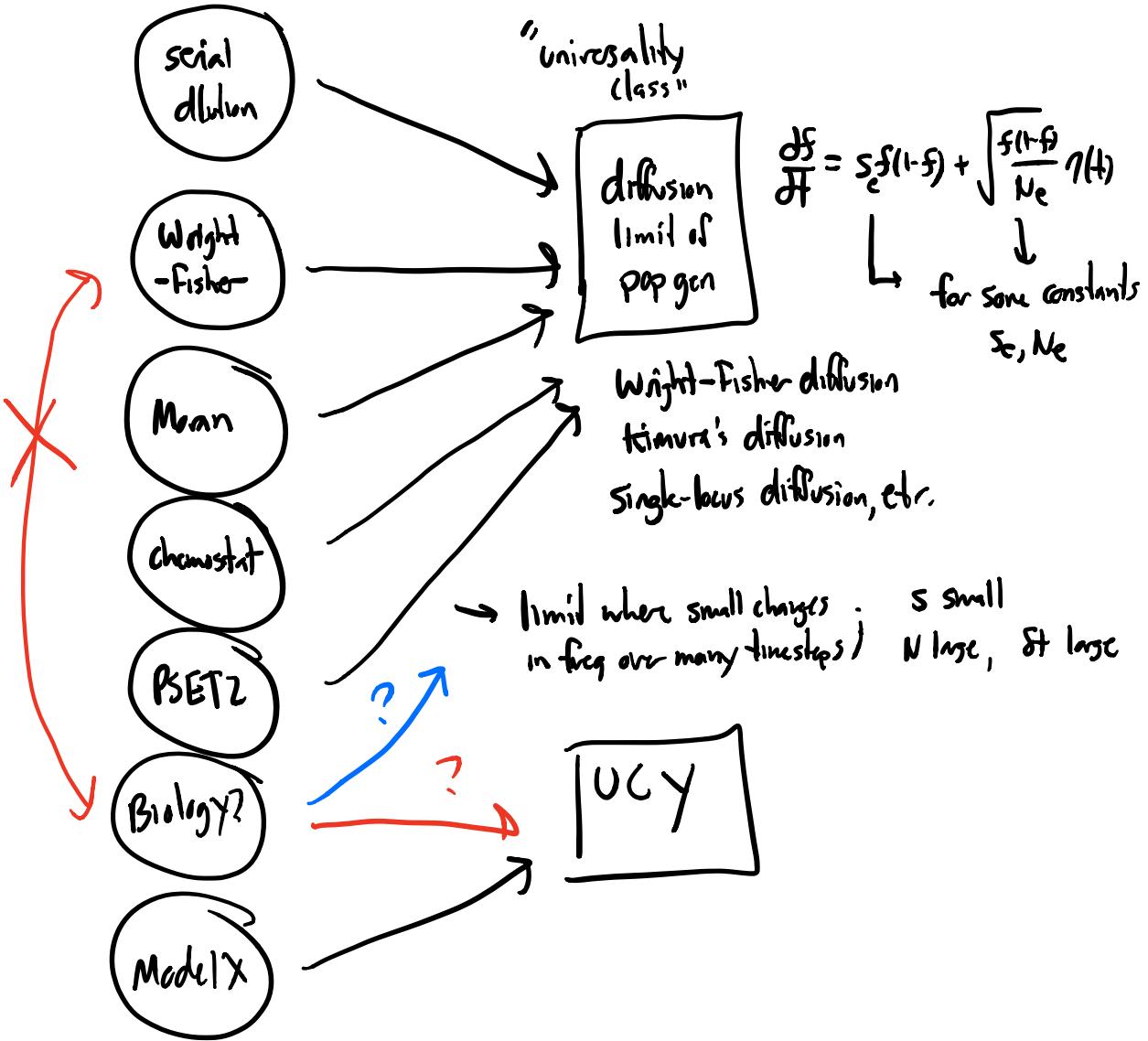
↙ when this applies:

$$f(t + \delta t) \approx f(t) + sf(1-f) \delta t + \sqrt{\frac{sf(1-f)}{N_e} \delta t} Z_+$$

"diffusion limit" of
population genetics
use as shorthand

$$\text{SDE: } \frac{df}{dt} = sf(1-f) + \sqrt{\frac{sf(1-f)}{N_e}} \eta(t)$$

"microscopic models"



"Traditional derivation" of diffusion limit of pop gen:

For arbitrary markov process:

$$x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_t \quad \text{w/} \quad \Pr(x_i \rightarrow x_{i+1} | x_i) \equiv p_i(x_{i+1} | x_i)$$



$p(x, t | x_0) = \text{prob of being at position } x \text{ at time } t \text{ given } x_0$

\Rightarrow then consider all positions at previous timestep!

$$\begin{aligned} p(x, t+1 | x_0) &= \int dx' p(x', t | x_0) p_i(x | x') \xrightarrow{\substack{\text{recursive} \\ \text{formula for} \\ p(x, t | x_0)}} \\ &= \int dx \ p(x - \Delta x, t | x_0) p_i(x | x - \Delta x) \end{aligned}$$

\Rightarrow Taylor expand in time & Δx

$$p(x, t+1 | x_0) \approx p(x, t | x_0) + \partial_t p(x, t | x_0)$$

$$p(x - \Delta x, t | x_0) \approx p(x, t | x_0) - \Delta x \partial_x p(x, t | x_0) + \frac{1}{2} \Delta x^2 \partial_x^2 p(x, t | x_0)$$

↳ collecting terms, can write as → Ewens for derivation

$$\frac{\partial \rho(x,t|x_0)}{\partial t} = -\frac{\partial}{\partial x} \left[\nu(x) \rho(x,t|x_0) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[\sigma^2(x) \rho(x,t|x_0) \right]$$

where $\nu(x) = \int \Delta x \rho_1(x+\Delta x|x)$

$$\sigma^2(x) = \int \Delta x^2 \rho_1(x+\Delta x|x)$$

"Fokker-Planck" equation
or "forward equation"

↑ equivalent to SDE (Langevin eq)

$$\frac{dx}{dt} = \nu(x) + \sqrt{\sigma^2(x)} \eta(t)$$

↑ or recursive formula:

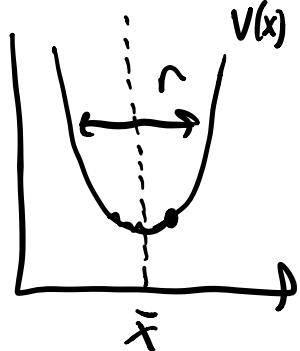
$$x(t+\delta t) = \nu(x) \delta t + \sqrt{\sigma^2(x) \delta t} Z_+$$

Manipulating SDEs:

① classic example: Brownian particle in quadratic potential

$$\frac{dx}{dt} = -\overbrace{r(x-\bar{x})}^{\text{restoring force}} + \overbrace{\sqrt{D}\eta(t)}^{\text{diffusion constant}}$$

\bar{x} equilibrium point



e.g. say we're interested in mean $\langle x(t) \rangle$

from definition: $x(t+\delta t) = x(t) - r(x-\bar{x})\delta t + \sqrt{D\delta t} \tilde{z}_+$ $\sim N(0,1)$

$$\langle x(t+\delta t) \rangle = \langle x(t) \rangle - r\langle x(t) \rangle \delta t + r\bar{x}\delta t + \circlearrowright$$

$$\Downarrow \quad \frac{\langle x(t+\delta t) \rangle - \langle x(t) \rangle}{\delta t} = -r[\langle x(t) \rangle - \bar{x}]$$

\Downarrow

$$\frac{d\langle x \rangle}{dt} = -r[\langle x \rangle - \bar{x}] \quad \text{ODE for } \langle x \rangle$$

$$\Rightarrow \langle x(t) \rangle - \bar{x} = (x(0) - \bar{x}) e^{-rt}$$

" $\langle x(t) \rangle \rightarrow \bar{x}$ @ rate r "