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

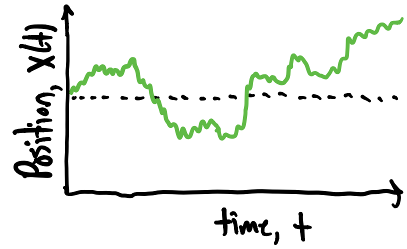
Last time: Random walks & diffusion

$$\underbrace{\Delta x_1 + \Delta x_2 + \Delta x_3 +}\ \cdots\ \underbrace{+}\ \underbrace{\Delta x_{t-2} + \Delta x_{t-1} + \Delta x_t}$$

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

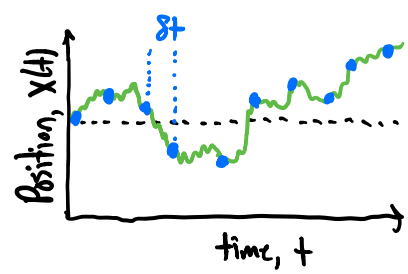
$$x(t+1) = \Delta x_1 + \Delta x_2 + \Delta x_3 + \dots + \Delta x_{t-2} + \Delta x_{t-1} + \Delta x_t$$

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



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

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



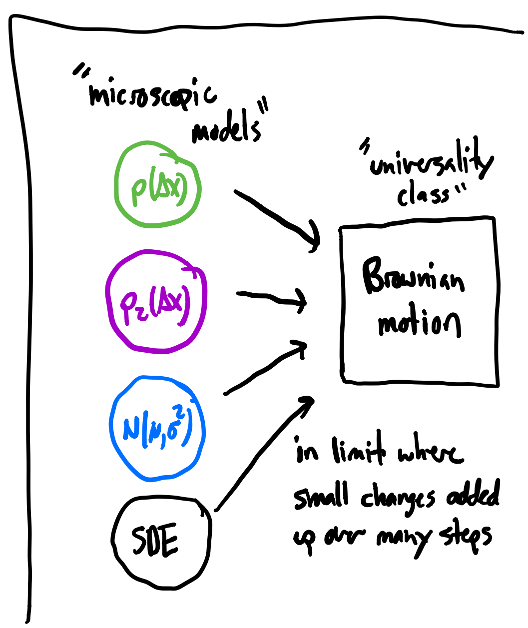
$\Downarrow$ will write as

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

or

in this class, shorthand for

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



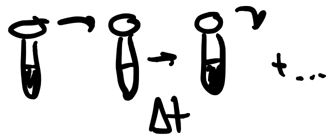
Today: ① Can this apply to evolution?

② working w/ SDEs.

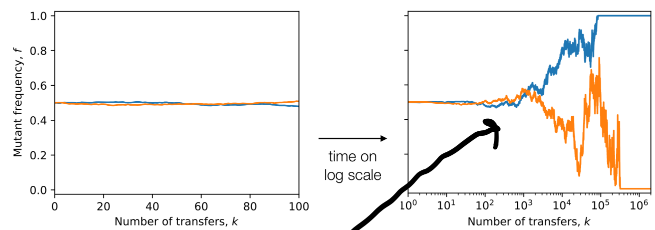
Basic idea: evolutionary phenomena take place over many generations...

$\Rightarrow$ maybe similar diffusion-like behavior emerges when $1 \ll \delta t \ll t$?

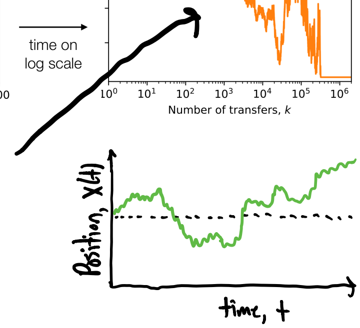
E.g. serial dilution model:



Computer simulations of model with $s = 0$, $N_0 = 10^5$, $f(0) = 50\%$



mutation frequency looks a bit like random walk...



$\Rightarrow$ can we use same approach here?

$$f(t+\Delta t) = f_0 + \underbrace{\Delta f_1 + \Delta f_2 + \Delta f_3}_{\downarrow} + \underbrace{\ldots}_{\downarrow} + \underbrace{\Delta f_{k-2} + \Delta f_{k-1} + \Delta f_k}_{\downarrow} \qquad \left(k = \frac{t}{\Delta t}\right)$$

coarse grain over timescale $\delta t \gg \Delta t$

$$= f_0 + \delta f_1 + \ldots + \delta f_{(t/\delta t)}$$

in this case:

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

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

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

$$\delta f_1 = \Delta f_1 + \Delta f_2 + \dots + \Delta f_{\left(\frac{\delta t}{\Delta t}\right)}$$ → can make this # large ~~→~~ can't apply CLT from before

→ $\mu(f(t))$
$\sigma^2(f(t))$

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

<u>Key idea</u>: if we coarse grain over many gens (for CLT) [lower bound on δt]

* but sufficiently few gens s.t. $f(t+\delta t) \simeq f(t)$ [upper bound]

then $\mu(f(t+i\Delta t)) \simeq \mu(f(t))$ for all $i \leq \delta t/\Delta t$

σ^2 ...

⇒ can therefore apply CLT from before:

$$\delta f = \Delta f_1 + \Delta f_2 + \dots + \Delta f_{\left(\frac{\delta t}{\Delta t}\right)} = \text{Gaussian}\left(\mu(f(t)) \cdot \overset{k = \left(\frac{\delta t}{\Delta t}\right)}{\downarrow}, \sigma^2(f(t))\delta t\right)$$

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

⇒ can we show that this works? (& <u>when</u> it works?)

⇒ use self-consistency & series approximations. * HW Problem

Step 1: what limits are relevant? (e.g. N_0, s, 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 $\boxed{s\Delta t \ll 1}$

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

$\xRightarrow[\text{large } N_0]{} \quad N_0 f \pm \sqrt{N_0 f} \Rightarrow f(t+\Delta t) \approx f \pm \sqrt{\frac{f}{N_0}}$

$\Rightarrow$ will be true when $\boxed{N_0 \gg 1}$

(strictly speaking, $N_0 f \gg 1$, $N_0 f(1-f) \ll 1$)

$\Rightarrow$ will discuss later.

<u>Step 2</u>: calculate leading order contributions to

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

in single timestep (using limits above)

① argument of Poisson: $\dfrac{f(t) e^{s\Delta t}}{f(t) e^{s\Delta t} + 1 - f(t)} \approx \dfrac{f(t)[1 + s\Delta t + \ldots]}{f(t)[1+s\Delta t + \ldots] + 1 - f}$

$$\simeq f(1+s\Delta t)(1 - f s\Delta t)$$

$$\simeq f + \underbrace{s\Delta t\, f(1-f)}_{\text{leading order contribution}} + \text{h.o.t.}$$

② Gaussian approx to Poisson dist'n:

$$f(t+\Delta t) = \frac{N_0[f + s\Delta t f(1-f)] + \sqrt{N_0 f}\, Z_1}{(\cdots) + N_0[1 - f - s\Delta t f(1-f)] + \sqrt{N_0(1-f)}\, Z_2}$$

$$= \frac{f + s\Delta t f(1-f) + \sqrt{\frac{f}{N_0}}\, Z_1}{1 + \sqrt{\frac{f}{N_0}}\, Z_1 + \sqrt{\frac{1-f}{N_0}}\, Z_2}$$

$$\approx f + s\Delta t f(1-f) + \underbrace{\sqrt{\frac{f}{N_0}}z_1 - f\sqrt{\frac{f}{N_0}}z_1}_{\displaystyle \underbrace{\sqrt{\frac{(1-f)^2 f}{N_0}}z_1 - \sqrt{\frac{f^2(1-f)}{N_0}}z_2}_{\displaystyle \sqrt{\underbrace{\frac{(1-f)^2 f}{N_0} + \frac{f^2(1-f)}{N_0}}} \, z_3}} - f\sqrt{\frac{1-f}{N_0}}z_2 + \text{h.o.t.}$$

$$\boxed{f(t+\Delta t) \approx f + \underbrace{s\Delta t f(1-f)}_{\mu(f)} + \sqrt{\frac{f(1-f)}{\underbrace{N_0}_{\Downarrow \, \sigma^2(f)}}}\, z} \quad \begin{pmatrix} s\Delta t \ll 1 \\ N_0 \gg \text{big} \end{pmatrix}$$

(that's one cycle)

<u>Step 3</u>: add up contributions over δt gens $\left(\frac{\delta t}{\Delta t} \text{ cycles}\right)$

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

$$= \underbrace{s}_{\Downarrow} f(1-f)\delta t + \sqrt{\frac{f(1-f)\delta t}{\underbrace{(N_0 \Delta t)}}}\, z_+$$

"s_e" effective selection strength

"N_e" effective population size.

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

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

   $\Rightarrow$ (a) Need $s\delta t \ll 1 \Rightarrow \delta t \ll 1/s$ (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

$$\boxed{\begin{array}{l} s \to 0 \\ N_0 \to \infty \\ \text{while } N_0 s \Delta t \text{ can be anything.} \end{array}}$$

✓

$\Downarrow$ when this applies:

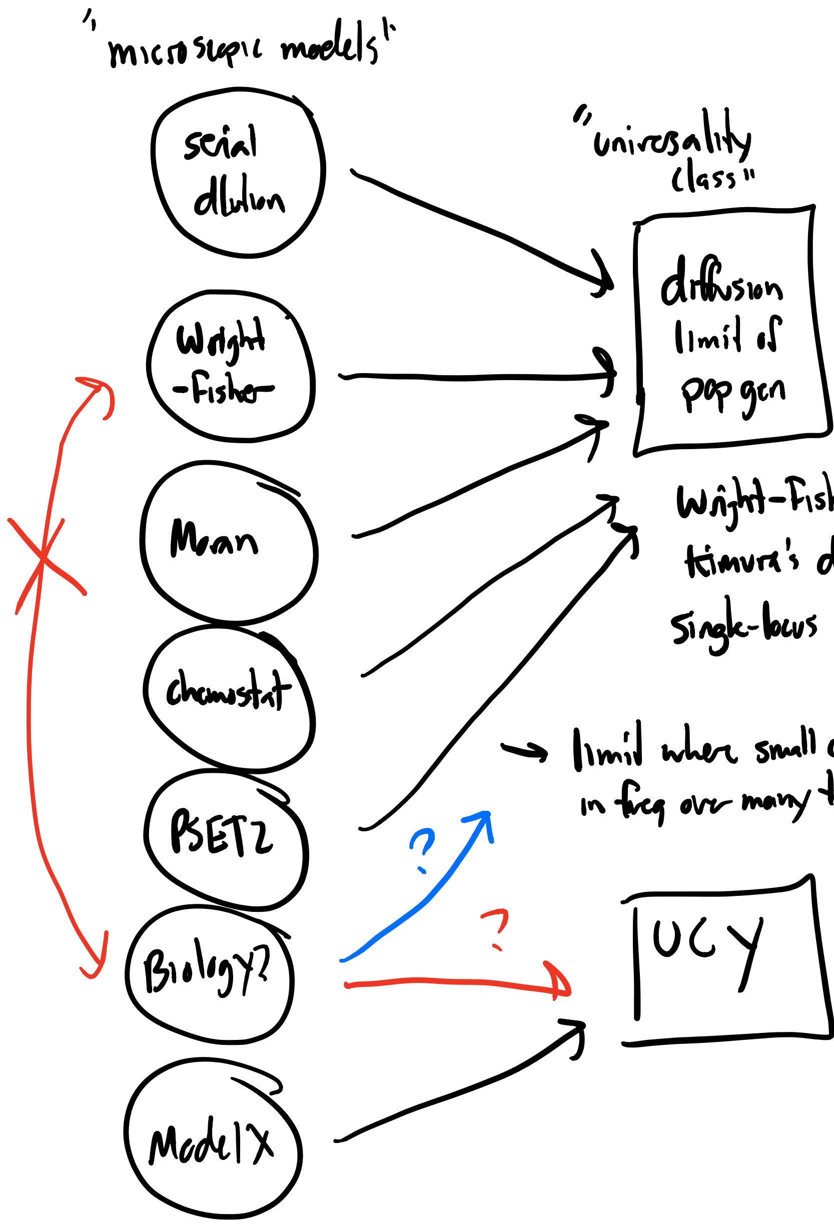
$$f(t+\delta t) \approx f(t) + sf(1-f)\,\delta t + \sqrt{\frac{f(1-f)\,\delta t}{N_e}}\, Z_t$$

"diffusion limit" of population genetics

use as shorthand

$\Updownarrow$

SDE: $$\frac{\partial f}{\partial t} = sf(1-f) + \sqrt{\frac{f(1-f)}{N_e}}\, \eta(t)$$



$$\frac{df}{dt} = s_e f(1-f) + \sqrt{\frac{f(1-f)}{N_e}}\,\eta(t)$$

for some constants s_e, N_e

Wright-Fisher diffusion
Kimura's diffusion
Single-locus diffusion, etc.

→ limit where small changes in freq over many timesteps: s small, N large, δt large

"Traditional derivation" of diffusion limit of pop gen:

For arbitrary markov process:

$$X_0 \to x_1 \to x_2 \to \dots \to X_t \quad w/ \quad Pr(x_i \to x_{i+1} | X_i) \equiv p_1(x_{i+1}|x_i)$$

$\Downarrow$

$p(x,t|x_0)$ = prob of being @ position x @ time t given $x=x_0$ @ $t=0$

$\Rightarrow$ then consider all positions @ previous timestep!

$$p(x,t+1|x_0) = \int dx' \; p(x',t|x_0)\, p_1(x|x') \quad \nearrow \text{recursive formula for } p(x,t|x_0)$$

$$= \int d\Delta x \; p(x-\Delta x, t|x_0)\, p_1(x|x-\Delta x)$$

$\Rightarrow$ Taylor expand in time & Δx

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

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

↳ collecting terms, can write as → $\boxed{\text{Ewens for derivation}}$

$$\frac{\partial p(x,t|x_0)}{\partial t} = -\frac{\partial}{\partial x}\left[N(x)\,p(x,t|x_0)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(f)\,p(x,t|f_0)\right]$$

where $N(x) = \int \Delta x \, p_1(x+\Delta x|x)$

$\sigma^2(\Delta x) = \int \Delta x^2 \, p_1(x+\Delta x|x)$

"Fokker-Planck" equation
-or- "forward equation"

⇕ equivalent to SDE (Langevin eq)

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

⇕ or recursive formula:

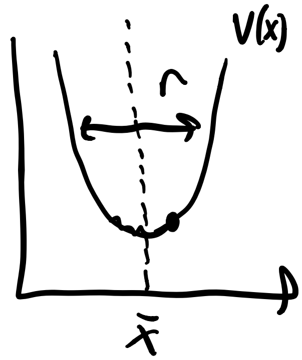
$$x(t+\delta t) = N(x)\,\delta t + \sqrt{\sigma^2(x)\,\delta t}\;Z_t$$

Manipulating SDEs:

① classic example: Brownian particle in quadratic potential

$$\frac{dx}{dt} = -\overbrace{r(x-\bar{x})}^{-\frac{dV}{dx}} + \sqrt{D}\,\eta(t)$$

restoring force — equilibrium point $\bar{x}$ — diffusion constant



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

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

$$\langle x(t+\delta t) \rangle = \langle x(t) \rangle - r\langle x(t) \rangle \delta t + r\bar{x}\delta t + 0$$

$$\Rightarrow \frac{\langle x(t+\delta t) \rangle - \langle x(t) \rangle}{\delta t} = -r\left[\langle x(t) \rangle - \bar{x}\right]$$

$$\Downarrow$$

$$\frac{\partial \langle x \rangle}{\partial t} = -r\left[\langle x \rangle - \bar{x}\right] \qquad \text{ODE for } \langle x \rangle$$

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

" $\langle x(t) \rangle \to \breve{x}$ @ rate r "

<u>Supplement</u>: "traditional" derivation of Fokker-Planck equation

Start w/ arbitrary Markov process: $x_0 \to x_1 \to \ldots \to x_t$

w/ single-step transition probability: $\Pr[x_{t+1}|x_t] \equiv p_1(x_{t+1}|x_t)$

Consider $p(x,t|x_0) \equiv$ prob density of being @ position x @ time t given that $x=x_0$ @ $t=0$.

By summing over positions @ previous time step, this probability satisfies recursion relation:

$$p(x,t+1) = \int dx' \, p(x',t|x_0) \times p_1(x|x')$$

switch from x' to $\Delta x = x - x'$

$$= \int d\Delta x \; p(x-\Delta x, t|x_0) \, p_1(x|x-\Delta x)$$

Useful to consider <u>generating function</u> $H(z) \equiv \int e^{-zx} p(x,t|x_0)\,dx$

(previously argued that $H(z) \Leftrightarrow p(x,t|x_0)$)

By integrating our recursion relation, we have:

$$H(z,t+1) \equiv \int dx\, e^{-zx}\, p(x,t+1|x_0) = \int dx\, d\Delta x\, e^{-zx}\, p(x-\Delta x,t|x_0)\, p_1(x|x-\Delta x)$$

$$\left(\begin{array}{l}\text{change vars to}\\ \tilde{x} = x-\Delta x\end{array}\right) \qquad = \int d\tilde{x}\, d\Delta x\, e^{-z\tilde{x}}\, e^{-z\Delta x}\, p(\tilde{x},t|x_0)\, p_1(\tilde{x}+\Delta x|\tilde{x})$$

$$(\text{relabel } \tilde{x} \to x) \qquad = \int dx\, d\Delta x\, e^{-zx}\, e^{-z\Delta x}\, p(x,t|x_0)\, p_1(x+\Delta x|x)$$

Now Taylor expand for small Δx:

$$\approx \int dx\, d\Delta x \cdot e^{-zx} \cdot \left[1 - z\Delta x + \tfrac{1}{2}(z\Delta x)^2\right] p(x,t|x_0)\, p_1(x+\Delta x|x)$$

$$= \int dx\, e^{-zx}\, p(x,t|x_0) \int d\Delta x\, p_1(x+\Delta x|x)\left[1 - z\Delta x + \tfrac{1}{2}(z\Delta x)^2\right]$$

$$= \int dx\, e^{-zx}\, p(x,t|x_0)\left[1 - z\mu(x) + \tfrac{1}{2}z^2\sigma^2(x)\right]$$

$$\sigma^2(x) \equiv \int \Delta x^2\, p_1(x+\Delta x|x)\, d\Delta x$$

$$\mu(x) \equiv \int \Delta x\, p_1(x+\Delta x|x)\, d\Delta x$$

$\Downarrow$ integration by parts

$$= \int dx\, e^{-zx}\left\{p(x,t|x_0) - \frac{\partial}{\partial x}\left[\mu(x)\, p(x,t|x_0)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x)\, p(x,t|x_0)\right]\right\}$$

thus, if $p(x,t|x_0)$ satisfies

$$\underbrace{p(x,t+1|x_0) - p(x,t|x_0)}_{\approx \frac{\partial p(x,t|x_0)}{\partial t}} = -\frac{\partial}{\partial x}\left[\mu(x)\,p(x,t|x_0)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x)\,p(x,t|x_0)\right]$$

$\Rightarrow$ then $H(z,t)$ [& $p(x,t|x_0)$] satisfy the recursion relation.