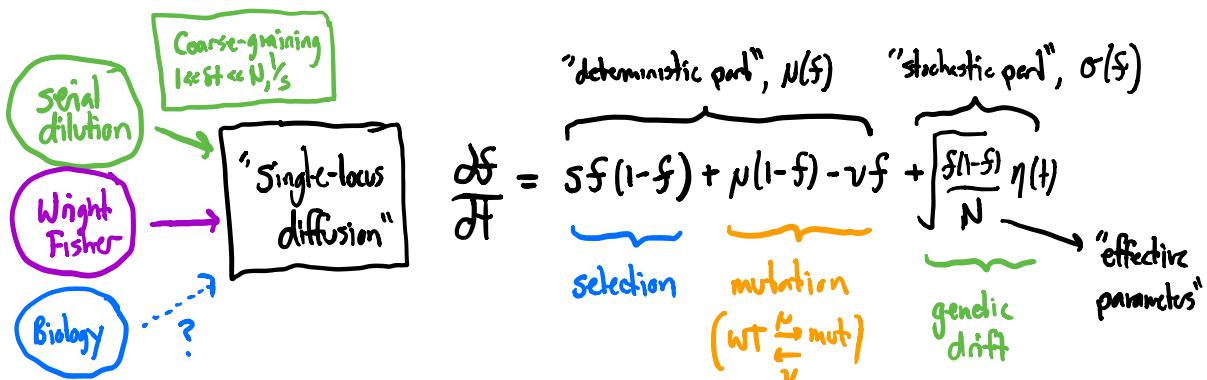


Announcements: Solutions for Problem Set 1 posted on Slack

Last time:



- Plan:
- ① Can we understand this model mathematically? (-4 lectures)
 - ② Back to reality: DNA sequencing & genomics
 - ③ Multi-site genomes ...

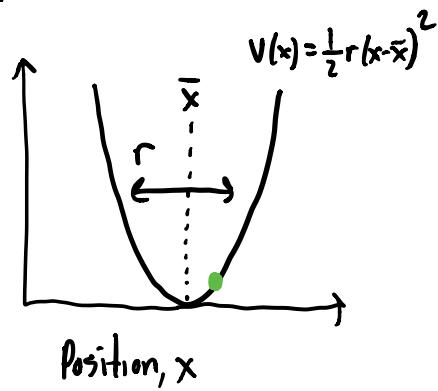
Last time: learn about SDEs w/ classic example:

"Brownian particle in a quadratic potential"

$$\frac{dx}{dt} = -\frac{\partial V(x)}{\partial x} + \sqrt{D} \eta(t)$$

Annotations for the equation:

- restoring force: $-\frac{\partial V(x)}{\partial x}$
- equilibrium point: \bar{x}
- diffusion constant: $(\alpha kT \text{ in physics})$



w/o noise ($D=0$), deterministic solution is $x_{\text{det}}(t) = x(0)e^{-rt} + \bar{x}(1-e^{-rt})$
 (approaches \bar{x} @ rate r)

with noise?

\Rightarrow can focus on moments, e.g. mean $\langle x(t) \rangle \xrightarrow{\sim} N(0,1)$

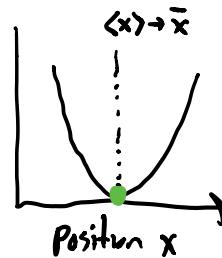
$$\begin{aligned}\langle x(t+\delta t) \rangle &= \langle x(t) - r[x(t)-\bar{x}] \delta t + \sqrt{D \delta t} Z_t \rangle \quad (\text{from "definition" of SDE}) \\ &= \langle x(t) \rangle - r[\langle x(t) \rangle - \bar{x}] \delta t + \sqrt{D \delta t} \langle Z_t \rangle^0\end{aligned}$$

When δt is small:

$$\Rightarrow \frac{\langle x(t+\delta t) - x(t) \rangle}{\delta t} \approx \boxed{\frac{d\langle x(t) \rangle}{dt} = -r[\langle x(t) \rangle - \bar{x}]} \quad (\text{ODE for } \langle x(t) \rangle)$$

\Rightarrow same as deterministic solution,

$$\langle x(t) \rangle = x(0)e^{-rt} + \bar{x}(1-e^{-rt})$$



What about spread around this value?

\Rightarrow can look @ higher moments, e.g. if $\bar{x}=0$, want $\langle x(t)^2 \rangle$

can use same basic idea:

Step 1:

$$\langle x(t+\delta t)^2 \rangle = \left\langle \left[x(t) - r x(t) \delta t + \sqrt{D \delta t} Z_t \right]^2 \right\rangle \quad \begin{matrix} \text{(from "definition")} \\ \text{of } x(t+\delta t) \end{matrix}$$

Step 2: expand to leading order in δt

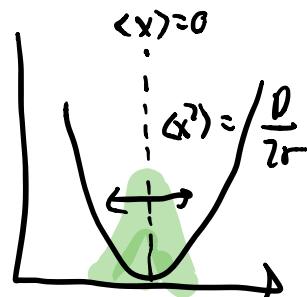
$$\begin{aligned} \langle x(t+\delta t)^2 \rangle &= \langle x(t)^2 - 2r x(t) \delta t \cdot x(t) + D \delta t Z_t^2 + 2x \sqrt{D \delta t} Z_t + \dots \rangle \\ &= \langle x(t)^2 \rangle - 2r \langle x(t)^2 \rangle \delta t + D \delta t \underbrace{\langle Z_t^2 \rangle}_1 + 2 \langle x \rangle \sqrt{D \delta t} \underbrace{\langle Z_t \rangle}_0 \end{aligned}$$

Step 3: take limit that δt small (reorganize)

$$\frac{\langle x(t+\delta t)^2 \rangle - \langle x(t)^2 \rangle}{\delta t} \approx \frac{d\langle x(t)^2 \rangle}{dt} = \underbrace{-2r \langle x^2 \rangle}_{\substack{\text{same as} \\ \text{deterministic} \\ \text{version}}} + \underbrace{D}_{\substack{\uparrow \text{new part from} \\ \text{stochastic part.}}}$$

$$\Rightarrow \frac{d\langle x^2 \rangle}{dt} = 0 \Rightarrow \langle x^2 \rangle = \frac{D}{2r}$$

"balance between
noise & deterministic
restoring force."



can actually get full dist'n @ long times:

$$\frac{dx}{dt} = -\frac{\partial V(x)}{\partial x} + \sqrt{D} \eta(t) \Leftrightarrow \frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x} \left[\frac{-\partial V}{\partial x} p \right] + \frac{D}{2} \frac{\partial^2}{\partial x^2} \left[D p \right]$$

("Fokker-Planck eq")

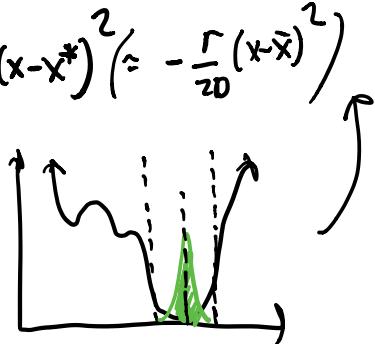
$$@ \text{ long times, } \frac{\partial p}{\partial t} = 0 \Rightarrow 0 = -\frac{\partial}{\partial x} \left[\frac{-\partial V}{\partial x} p(x) \right] + \frac{D}{2} \frac{\partial^2 p}{\partial x^2}$$

$$\Rightarrow -\frac{\partial V}{\partial x} p = \frac{D}{2} \frac{\partial p}{\partial x} \Rightarrow \frac{\partial \log p}{\partial x} = \frac{1}{D} \frac{\partial V}{\partial x} \quad \begin{matrix} \text{(Gaussian dist'n} \\ \text{w/ mean } \bar{x} \end{matrix}$$

$$\Rightarrow \log p(x) = -\frac{V(x)}{D} \quad \begin{matrix} \nearrow \\ \text{variance } \propto D \end{matrix}$$

$$\Rightarrow p(x) \propto e^{-\frac{V(x)}{D}} \propto e^{-\frac{1}{2D} \frac{\partial^2 V}{\partial x^2} (x-x^*)^2} \quad \begin{matrix} \left(\approx -\frac{1}{2D} (x-\bar{x})^2 \right) \end{matrix}$$

"Boltzmann distribution"



What about our evolutionary model?

e.g. $\frac{df}{dt} = sf(1-f) + N(1-f) - \nu f + \sqrt{\frac{f(1-f)}{N}} \eta(t)$ "It's famous
Stochastic framework"

2 key differences: ① $D_{\text{eff}} \approx \frac{f(1-f)}{N}$ (depends on mutation freq!)

② selection term ($sf(1-f)$) is non-linear

e.g. focus on $\langle f(t) \rangle$

Step 1:

$$\begin{aligned}\langle f(t+\delta t) \rangle &= \langle f(t) + \delta t [sf(1-f) + N(1-f) - \nu f] + \sqrt{\frac{f(1-f)}{N} \delta t} Z_t \rangle \\ &= \langle f(t) \rangle + \delta t [s(\langle f \rangle - \langle f^2 \rangle) + N(1 - \langle f \rangle) - \nu \langle f \rangle] + 0\end{aligned}$$

$$\Rightarrow \frac{d\langle f(t) \rangle}{dt} = s[\langle f \rangle - \langle f^2 \rangle] + N(1 - \langle f \rangle) - \nu \langle f \rangle$$

\Downarrow

$\not\equiv \langle f \rangle^2 \Rightarrow$ need $\langle f^2(t) \rangle$
to find $\langle f(t) \rangle$

do same thing for $\langle f(t+\delta t)^2 \rangle$

$$\dots \Rightarrow \frac{d\langle f^2 \rangle}{dt} = \underbrace{2s\langle f \cdot f(1-f) \rangle}_{\text{from deterministic part.}} + \underbrace{\frac{\langle f(1-f) \rangle}{N}}_{\text{from 2 stochastic terms, } \langle Z_f^2 \rangle = 1} + \text{mut'n's.}$$

\Rightarrow depends on $\langle f^3 \rangle$ in addition to $\langle f \rangle, \langle f^2 \rangle$

\Rightarrow known as "moment bell" (general consequence of nonlinearity)

\Rightarrow one solution: focus on $s=0$ ("neutral theory")

What about the stationary distribution?

Fokker-Plank equation for evolutionary model:

$$(\text{remember: } \frac{dx}{dt} = N(x) + \sqrt{\sigma^2(x)} \eta(t) \Leftrightarrow \frac{dp}{dt} = -\frac{d}{dx} [N(x)p] + \frac{1}{2} \frac{d^2}{dx^2} [\sigma^2(x)p])$$

(1) here:

$$\frac{dp}{dt} = -\frac{d}{df} \left[\left[sf(1-f) + N(1-f) - vf \right] p \right] + \frac{1}{2} \frac{d^2}{df^2} \left(\frac{f(1-f)p}{N} \right)$$

\Rightarrow when $\partial_t p \approx 0$ (equilibrium)

$$p(f) \propto f^{-1} (1-f)^{-1} e^{-2N\Lambda(f)}$$

where $\Lambda(f) = sf + \nu \log f + \nu \log(1-f)$

$$\Rightarrow \text{deterministically } \frac{d\Lambda}{dt} = \frac{\partial \Lambda}{\partial f} \frac{df}{dt} = -f(1-f) \left(\frac{df}{dt} \right)^2 \leq 0$$

\Rightarrow dynamics try to "minimize" $\Lambda(f)$ [like "energy" $V(x)$]

$\Rightarrow N$ is analogy of " $\frac{1}{\text{temp}}$ " (or "noise")

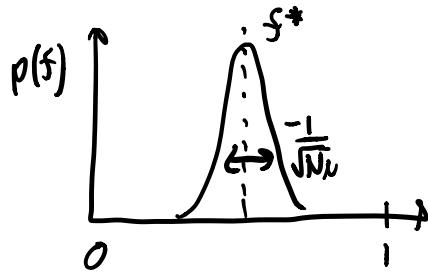
Plugging in for $\Lambda(f)$ in this example:

$$p(f) \propto f^{2N\mu-1} (1-f)^{2N\nu-1} e^{2Ns} \quad \begin{array}{l} \text{"mutation-selection-drift} \\ \text{balance"} \\ (\text{Wright 1930s}) \end{array}$$

what does this look like?

Case 1: $N\mu, N\nu \gg 1$

peaked @ minimum of $N(f^*)$



$$\frac{\partial N}{\partial f} = 0$$

"mutation-selection balance"

$$\Downarrow s + \frac{N}{f^*} - \frac{\nu}{1-f^*} = 0$$

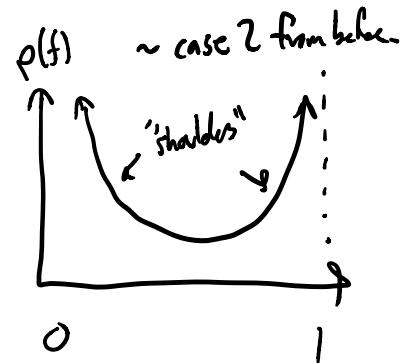
(same as deterministic solution for $\frac{df}{dt} = 0$)

in simple limits:

$$f^* \approx \begin{cases} \frac{\nu}{|s|} & \text{if } s < 0; |s| \gg \mu, \nu \\ \frac{\mu}{\mu+\nu} & \text{if } |s| \ll \mu, \nu \\ 1 - \frac{\nu}{s} & \text{if } s \gg \mu, \nu \end{cases}$$

full dist'n is $p(f) \propto (f^*)^{-1} (1-f^*)^{-1} e^{-N \frac{d^2 N(f^*)}{df^2} (f-f^*)^2}$

Case 2: $N\mu \ll 1, N\nu \ll 1 \Rightarrow$ "U-shaped form"



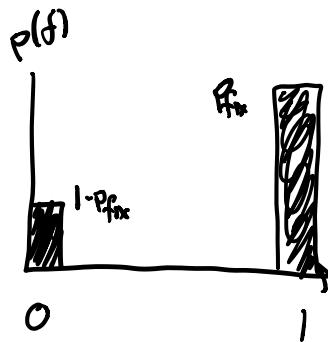
\Rightarrow where height of shoulders differs by e^{2Ns^2}

\Rightarrow definitely not deterministic + a little noise.

$$\left(\frac{df}{dt} = sf(1-f) + N(1-f) - \nu f + \sqrt{\frac{f(1-f)}{N}} \eta(t) \right)$$

one mac stationary dist'n scenario:

no mutations: $\frac{df}{dt} = sf(1-f) + \sqrt{\frac{f(1-f)}{N}} \eta(t)$



$$p(f) = p_{\text{fix}} \delta(f-1) + (1-p_{\text{fix}}) \delta(f)$$

where p_{fix} depend on f_0

$$p_{\text{fix}}(f_0) = \begin{cases} 0 & \text{if } f_0 = 0 \\ ? & \text{if } f_0 \text{ is } \text{fixed} \\ 1 & \text{if } f_0 = 1 \end{cases}$$

Fokker-Planck Eq. not helpful, but generating function is!

$$H(z,t) \equiv \langle e^{-zf(t)} \rangle = \int e^{-zf} p(f,t) df$$

↳ use same approach...

$$\begin{aligned} \langle H(z,t+\delta t) \rangle &= \langle e^{-z f(t+\delta t)} \rangle \\ &= \left\langle e^{-z \left[f(t) + s f(1-f) \delta t + \sqrt{\frac{f(1-f)}{N}} \delta t z \right]} \right\rangle \end{aligned}$$

= Taylor expand through $\Theta(\delta t) \approx \arg \text{avr } \bar{z}_t$

$$= \underbrace{\langle e^{-\bar{z}f(t)} \rangle}_{H(z,t)} + \left\langle e^{-\bar{z}f} \left[-\bar{z}sf(1-f) + \underbrace{\frac{\bar{z}^2}{2N} f(1-f)}_{\text{det part}} \right] \delta t \right\rangle + \underbrace{\bar{z} \text{ stochastic terms.}}_{\text{stochastic terms.}}$$

\Rightarrow rearrange:

$$\frac{H(z, t+\delta t) - H(z, t)}{\delta t} = \frac{\partial H(z, t)}{\partial t} = \left\langle - \left[sz - \frac{\bar{z}^2}{2N} \right] f(1-f) e^{-\bar{z}f} \right\rangle + \left[\left(-\frac{\partial}{\partial z} + \frac{\partial^2}{\partial z^2} \right) e^{-\bar{z}f} \right]$$

$$\Rightarrow \frac{\partial H}{\partial t} = \left[sz - \frac{\bar{z}^2}{2N} \right] \left[\frac{\partial H}{\partial z} - \frac{\partial^2 H}{\partial z^2} \right]$$

PDE for
 $H(z, t)$

hard in general, but for $\bar{z}^* = 2Ns$

$$\frac{dH(z^*, t)}{dt} = 0 \Rightarrow H(z^*, t) = \text{const.} = e^{-\bar{z}^* f_0}$$

$$H(z^*, \infty) = p_{fix} e^{-z^*} + (1-p_{fix}) e^{-0} = e^{-z^* f_0}$$

$$\Rightarrow P_{fix}(f_0) = \frac{1 - e^{-2Ns f_0}}{1 - e^{-2Ns}}$$

"fixation probability"
Kimura 1950's

$$\frac{df^2}{dt} = 2f \cdot \frac{df}{dt} = 2sf^2(1-f) + 2f\sqrt{\frac{f(1-f)}{N}}\eta(t)$$

$$\left\langle \frac{df^2}{dt} \right\rangle = \left\langle 2sf^2(1-f) \right\rangle + 0 \quad X$$

↑
+ $\left\langle \sqrt{\frac{f(1-f)}{N}}, \sqrt{\frac{f(1-f)}{N}} \cdot \eta \right\rangle^2$

$$\frac{dg(f)}{dt} = \left\langle \frac{dg}{df} \frac{df}{dt} \right\rangle + \left\langle \frac{1}{2} g''(f) \right\rangle \quad \text{"Ito formula"}$$