Announcements: Solutions for Problem Set 1 posted on Slack

Last time:

Serial dilution → Coarse-graining $1 \ll \delta t \ll N, 1/s$ → "Single-locus diffusion"

Wright Fisher → "Single-locus diffusion"

Biology ⇢ ?

$$\frac{\partial f}{\partial t} = \overbrace{\underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\mu(1-f) - \nu f}_{\text{mutation}}}^{\text{"deterministic part", } v(f)} + \overbrace{\underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}}^{\text{"stochastic part", } \sigma(f)}$$

mutation $\left(WT \underset{\nu}{\overset{\mu}{\rightleftharpoons}} mut\right)$

N → "effective parameters"

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:

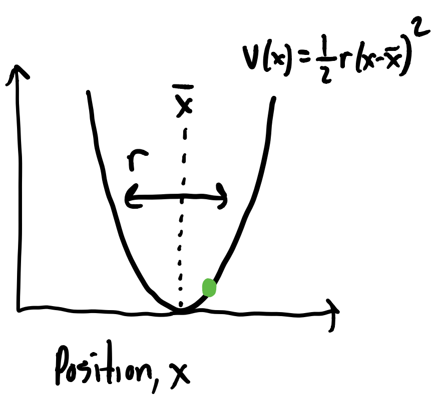
"Brownian particle in a quadratic potential"

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

r: restoring force

$\bar{x}$: equilibrium point

D: diffusion constant ($\propto kT$ in physics)



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

(approaches $\bar{x}$ @ rate r)

with noise?

$\Rightarrow$ can focus on moments, e.g. mean $\langle x(t) \rangle$

$$\langle x(t+\delta t) \rangle = \langle x(t) - r[x(t)-\bar{x}]\delta t + \sqrt{D\delta t}\, \underbrace{Z_t}_{N(0,1)} \rangle$$

(from "definition" of SDE)

$$= \langle x(t) \rangle - r[\langle x(t) \rangle - \bar{x}]\delta t + \sqrt{D\delta t}\, \cancel{\langle Z_t \rangle}^{0}$$

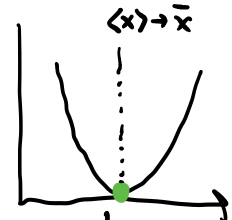
when δt is small:

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

(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:

$$\langle x(t+\delta t)^2 \rangle = \left\langle \left[x(t) - r\,x(t)\,\delta t + \sqrt{D\delta t}\, Z_t \right]^2 \right\rangle$$

(from "definition" of $x(t+\delta t)$)

Step 2: expand to leading order in δt

$$\langle x(t+\delta t)^2 \rangle = \langle x(t)^2 - 2r\,x(t)\,\delta t \cdot x(t) + D\delta t\, Z_t^2 + 2x\sqrt{D\delta t}\, Z_t + \dots \rangle$$

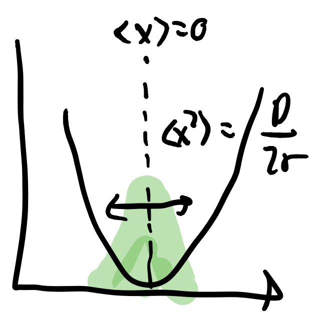
$$= \langle x(t)^2 \rangle - 2r\langle x(t)^2 \rangle \delta t + D\delta t \underbrace{\cancel{\langle Z_t^2 \rangle}}_{1} + 2\langle x \rangle \sqrt{D\delta t}\underbrace{\cancel{\langle Z_t \rangle}}_{0}$$

Step 3: take limit that δt small (reorganize)

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

$$\Rightarrow \frac{\partial \langle x^2 \rangle}{\partial t} \approx 0 \Rightarrow \langle x^2 \rangle = \frac{D}{2r}$$

"balance between noise & deterministic restoring force.



can actually get full dist'n @ long times:

$$\frac{dx}{dt} = -\frac{dV(x)}{dx} + \sqrt{D}\,\eta(t) \iff \frac{dp(x,t)}{dt} = \frac{-d}{dx}\left[\frac{-dV}{dx}\cdot p\right] + \frac{1}{2}\frac{d^2}{dx^2}\left[Dp\right]$$

("Fokker Planck eq")

@ long times, $\frac{dp}{dt}=0 \Rightarrow 0 = \frac{-d}{dx}\left[-\frac{dV}{dx}p(x)\right] + \frac{D}{2}\frac{d^2p}{dx^2}$

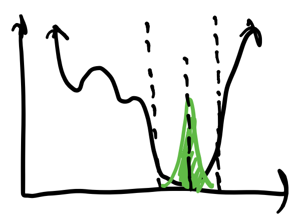
$$\Rightarrow -\frac{dV}{dx}p = D\frac{dp}{dx} \Rightarrow \frac{d\log p}{dx} = \frac{1}{D}\frac{-dV}{dx}.$$

$$\Rightarrow \log p(x) = -\frac{V(x)}{D} \Rightarrow$$

$$\Rightarrow p(x) \propto e^{-\frac{V(x)}{D}} \propto e^{-\frac{1}{2D}\frac{d^2V(x^*)}{dx^2}(x-x^*)^2} \left(\approx -\frac{r}{2D}(x-\bar{x})^2\right)$$

$\boxed{\text{"Boltzmann distribution"}}$

Gaussian dist'n w/ mean $\bar{x}$ & variance D



What about our evolutionary model?

e.g. $$\frac{\partial f}{\partial t} = sf(1-f) + \mu(1-f) - \nu f + \sqrt{\frac{f(1-f)}{N}}\,\eta(t)$$ "Ito formula stonovich formula"

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

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

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

Step 1:

$$\langle f(t+\delta t) \rangle = \Big\langle f(t) + \delta t\big[sf(1-f) + \mu(1-f) - \nu f\big] + \sqrt{\frac{f(1-f)\,\delta t}{N}}\; \overset{N(0,1)}{Z_t} \Big\rangle$$

$$= \langle f(t) \rangle + \delta t\big[s(\langle f \rangle - \langle f^2 \rangle) + \mu(1-\langle f \rangle) - \nu\langle f \rangle\big] + 0$$

$$\Rightarrow \quad \frac{\partial \langle f(t) \rangle}{\partial t} = s\big[\langle f \rangle - \langle f^2 \rangle\big] + \mu(1-\langle f \rangle) - \nu\langle f \rangle$$

$\Downarrow$

$\underline{\underline{\text{not}}}\ \langle f \rangle^2 \Rightarrow$ need $\langle f^2(t) \rangle$ to find $\langle f(t) \rangle$

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

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

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

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

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

$\boxed{\text{what about the stationary distribution?}}$

Fokker plank equation for evolutionary model:

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

(1) here:

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

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

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

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

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

$\Rightarrow$ 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:

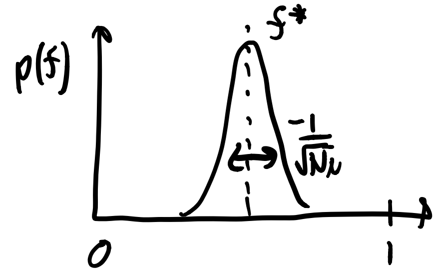
$$\boxed{p(f) \propto f^{2N\mu-1}(1-f)^{2N\nu-1} e^{2Nsf}}$$

"mutation-selection-drift balance"
(Wright 1930s)

what does this look like?

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

peaked @ minimum of $\Lambda(f^*)$



$\frac{\partial \Lambda}{\partial f} = 0$ $\boxed{\text{"mutation-selection balance"}}$

$\Rightarrow s + \frac{\mu}{f^*} - \frac{\nu}{1-f^*} = 0$

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

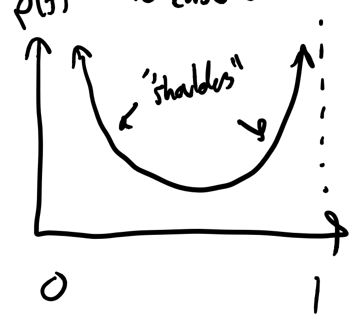
in simple limits:

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

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

~ case 2 from before

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

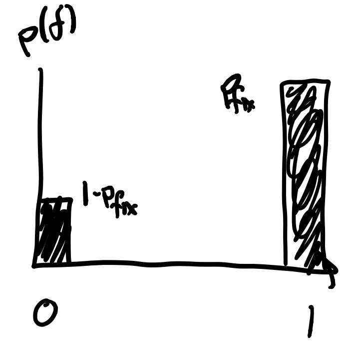


$\Rightarrow$ where height of shoulders differs by e^{2Nsf}

$\Rightarrow$ definitely not deterministic + a little noise.

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

one more stationary dist'n scenario:



no mutations : $\dfrac{\partial f}{\partial t} = sf(1-f) + \sqrt{\dfrac{f(1-f)}{N}}\,\eta(t)$

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

when p_{fix} depend on f_0

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

Fokker planck Eq. not helpful, but generating function is!

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

$\Rightarrow$ use same approach...

$$\langle H(z,t+\delta t) \rangle = \left\langle e^{-zf(t+\delta t)} \right\rangle$$

$$= \left\langle e^{-z\left[f(t) + sf(1-f)\delta t + \sqrt{\frac{f(1-f)}{N}\delta t}\, Z_t\right]} \right\rangle$$

= Taylor expand through $\mathcal{O}(\delta t)$ & avg over Z_t

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

$\Rightarrow$ rearrange:

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

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

PDE for $H(z,t)$

hard in general, but for $z^* = 2Ns$

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

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

$$\boxed{\Rightarrow \quad p_{fix}(f_0) = \frac{1-e^{-2Ns f_0}}{1-e^{-2Ns}} \qquad \text{"fixation probability"} \quad \text{Kimura 1950's}}$$