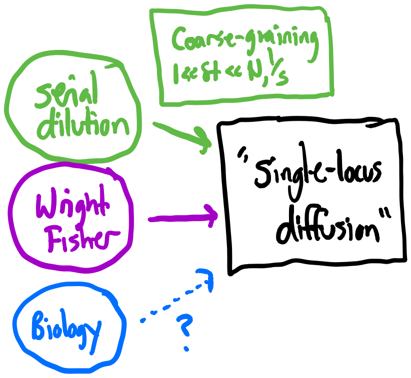
Announcements: Solutions for Problem Set 1 posted on Slack

Last time:



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

$\rightarrow$ "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:

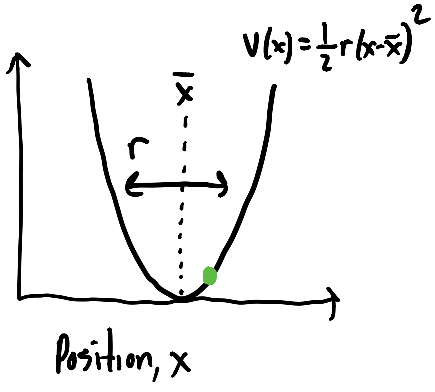
"Brownian particle in a quadratic potential"

$$\frac{dx}{dt} = \underbrace{-\overset{\substack{\text{restoring}\\\text{force}}}{r}(x-\overset{\substack{\text{equilibrium}\\\text{point}}}{\bar{x}})}_{-\frac{dV(x)}{dx}} + \sqrt{\overset{\substack{\text{diffusion}\\\text{constant}}}{D}}\,\eta(t)$$

restoring force

equilibrium point

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 $\langle x(t)\rangle$

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

(from "definition" of SDE)

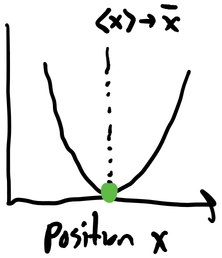
$$= \langle x(t)\rangle - r[\langle x(t)\rangle - \bar{x}]\delta t + \sqrt{D\delta t}\, \cancelto{0}{\langle Z_t\rangle}$$

when δt is small:

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

$\Rightarrow$ 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:

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

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

Step 2: expand to leading order in δt

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

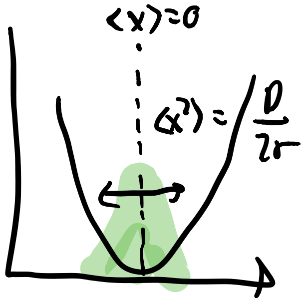
$$= \langle x(t)^2\rangle - 2r\langle x(t)^2\rangle\delta t + D\delta t\langle \underbrace{\cancel{Z_t^2}}_{1}\rangle + 2\langle x\rangle\sqrt{D\cancel{\delta t}}\underbrace{\langle Z_t\rangle}_{0}$$

Step 3: take limit that δt small (reorganize)

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

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

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

$$\Rightarrow -\frac{dV}{dx}p = \frac{D}{2}\frac{dp}{dx} \Rightarrow \frac{d\log p}{dx} = -\frac{2}{D}\frac{dV(x)}{dx}$$

$$\Rightarrow \log p(x) = -\frac{2V(x)}{D}$$

$$\Rightarrow p(x) \propto e^{-\frac{2V(x)}{D}} \propto e^{-\frac{1}{D}\frac{d^2V(x^*)}{dx^2}(x-x^*)^2} \xrightarrow{e.g.} -\frac{r}{D}(x-\bar{x})^2$$

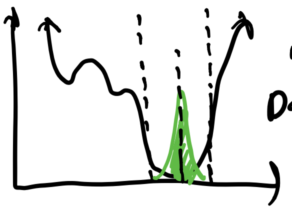
$\Uparrow$

Gaussian dist'n w/ mean $\bar{x}$ & variance $\frac{D}{2r}$

$\uparrow$

local approx when $D \ll V''(x^*)$

"Boltzmann distribution"



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

Stratonovich formula"

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:

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

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

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

$$\Downarrow$$

$\overset{\text{not}}{=} \langle f\rangle^2 \Rightarrow$ 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 $s \approx 0$ ("neutral theory")

what about the stationary distribution?

Fokker plank equation for evolutionary model:

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

$\Downarrow$ here:

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

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

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

where $\Lambda(f) = -\left[sf + \mu \log f + \nu \log(1-f)\right]$

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

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

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

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

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

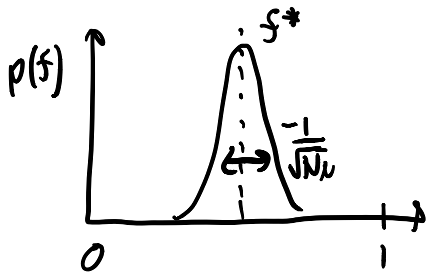
"mutation-selection-drift balance"

(Wright 1930s)

what does this look like?

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

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



$$\left.\frac{\partial\Lambda}{\partial f}\right|_{f=f^*} = 0 \qquad \boxed{\text{“mutation-selection balance”}}$$

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

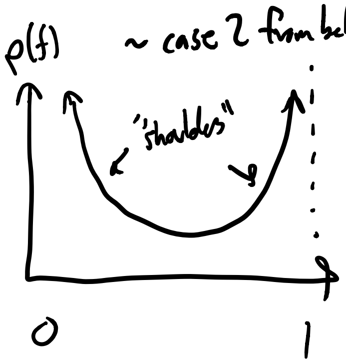
(same as deterministic sol'n, $\frac{df}{dt} = sf^*(1-f^*)+\mu(1-f^*)-\nu f^* = 0$)

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



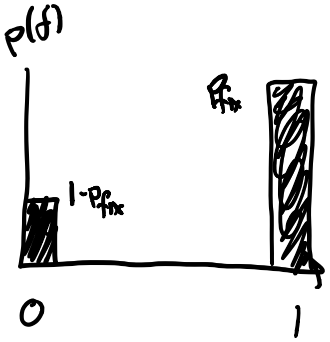
~ case 2 from before.

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

⇒ definitely not deterministic + a little noise.

$$\left(\frac{df}{dt} = sf(1-f)+\mu(1-f)-\nu f + \cancelto{0?\,\times}{\sqrt{\frac{f(1-f)}{N}}}\eta(t) \right)$$

one more stationary dist'n scenario:



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

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

where 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 plank 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...

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

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

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

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

$\Rightarrow$ 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 $z^* = 2Ns$

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

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

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

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

$$\left\langle \frac{\partial f^2}{\partial t} \right\rangle = \left\langle 2sf^2(1-f) \right\rangle + 0 \qquad X$$

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

$$\frac{\partial \cancel{f}\, g(f)}{\partial t} = \left\langle \frac{\partial g}{\partial f} \frac{\partial f}{\partial t} \right\rangle + \left\langle \frac{1}{2} g''(f) \right\rangle$$

"Ito formula"