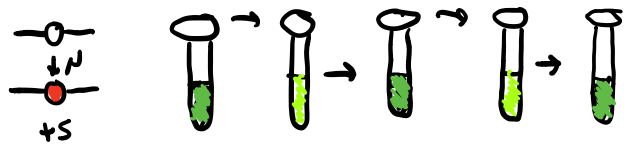
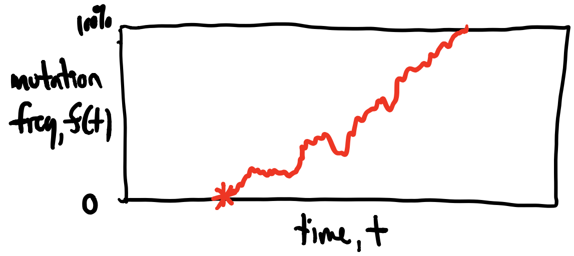
Announcements: PSET 2 updated to fix typo (thx Olivia + Xiran)

Last time: Quick review – how did we get here?



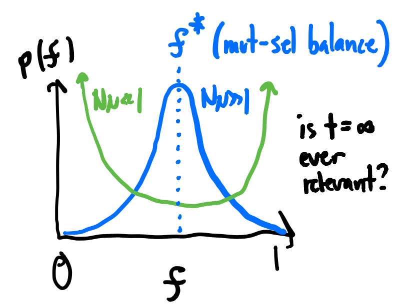
$$\Rightarrow \frac{\partial f}{\partial t} = \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}} + \underbrace{\left[\mu(1-f) - \nu f\right]}_{\text{mutation}}$$



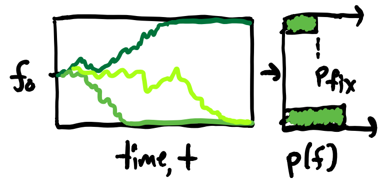
$$\frac{\partial p(f,t)}{\partial t} = -\frac{\partial}{\partial f}\left[\left(sf(1-f) + \mu(1-f) - \nu f\right) p(f,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial f^2}\left[\frac{f(1-f)}{N}\, p(f,t)\right]$$

① "Moment hell": $\frac{\partial \langle f\rangle}{\partial t} = s\langle f\rangle - \langle f^2\rangle$, etc...

② Stationary distribution $p(f,t) \xrightarrow[t\to\infty]{} \underbrace{p(f) \propto f^{2N\mu-1}(1-f)^{2N\nu-1}e^{2Nsf}}_{\text{"mutation-selection-drift balance"}}$

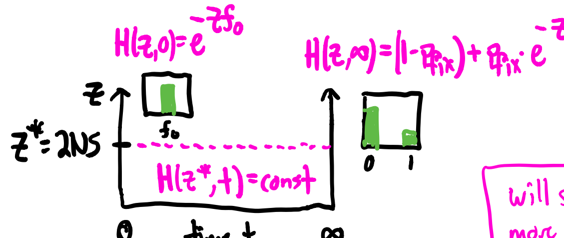


③ Fixation probability: (no mutation, $\mu=\nu=0$) $\quad p_{fix}(N,s,f_0) = \frac{1-e^{-2Nsf_0}}{1-e^{-2Ns}}$ "Kimura formula"



Generating function: $H(z,t) \equiv \langle e^{-zf(t)}\rangle$

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



will see more later!

Today: what can we learn from $p_{fix}(N,s,f_0) = \dfrac{1-e^{-2Nsf_0}}{1-e^{-2Ns}}$?

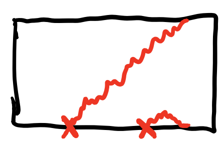
$\Rightarrow$ battle between selection & genetic drift $(N \times s)$

(a) If $N|s| \ll 1 \Rightarrow p_{fix} \approx f_0$ ("drift wins") ("weak selection" / "neutrality")

(b) If $N|s| \gg 1$ ("strong selection")

$$p_{fix}(N,s,f_0) \approx \begin{cases} 1 & \text{if } s>0;\ f_0 \gg 1/2Ns \longrightarrow \text{"selection wins"} \\ 2Nsf_0 & \text{if } s>0;\ f_0 \ll 1/2Ns \longrightarrow \text{outcome uncertain....} \\ e^{-2N|s|(1-f_0)} & \text{if } s<0; \longrightarrow \approx 0 \text{ "selection wins"} \end{cases}$$

e.g. extrapolate to new mutation $(f_0 = 1/N)$



$$\Rightarrow \boxed{p_{fix} \approx 2s \quad \text{(independent of } N!) \qquad \text{"Haldane's formula" (Haldane 1930s)}}$$

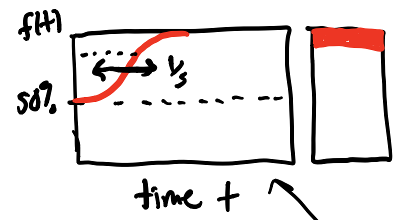
e.g. if $N \approx 10^7$, $s \approx 0.01$ (per gen.) $\Rightarrow$ only 2% chance that mutation fixes!

(pretty beneficial on lab timescales!)

Problem 4, Pset 2

$\Rightarrow$ 98% of these mutations go extinct due to genetic drift!

$\Rightarrow$ but same mutant mixed @ 50-50 will rapidly + consistently take over!



what's going on here?

naively, as $N \to \infty$: $\frac{\partial f}{\partial t} = sf(1-f) + \cancel{\sqrt{\frac{f(1-f)}{N}}}\,\eta(t)$ $\approx 0?$ $\Rightarrow f(t) = \frac{f(0)e^{st}}{f(0)e^{st} + 1 - f(0)}$

(deterministic solution)

$P_{fix} \approx \begin{cases} 1 & \text{if } s > 0 \\ 0 & \text{if } s < 0 \end{cases}$

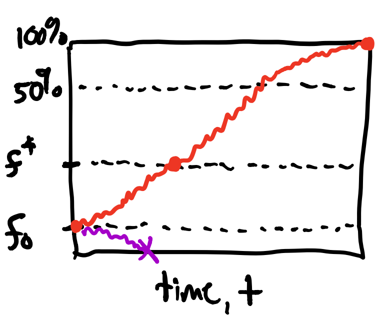
somehow wrong...

How can we understand this?

Note : $P_{fix} = \frac{1 - e^{-2Nsf_0}}{1 - e^{-2Ns}} \approx 1$ when $f_0 \gg \frac{1}{2Ns}$, even when $f_0 \ll 1$

$\Rightarrow$ i.e. outcome only uncertain when $f_0 \lesssim \frac{1}{2Ns}$ ($\ll 1$ when $Ns \gg 1$)

$\Rightarrow$ break into two parts : $f_0 \to f^*$ + $f^* \to 1$ (w/ $f^* \ll 1$)

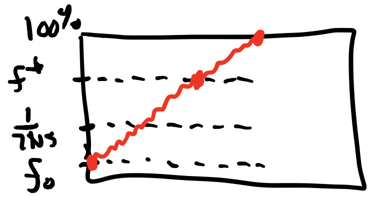


$\Rightarrow$ must have

$$P_{fix}(f_0) = Pr(f_0 \to f^*) \times P_{fix}(f^*)$$

$$\Rightarrow \Pr(f_0 \to f^*) = \frac{p_{fix}(f_0)}{p_{fix}(f^*)} \approx \frac{2Nsf_0}{p_{fix}(f^*)} \quad \text{if } f_0 \ll 1/2Ns$$

① if $f^* \gg 1/2Ns$ $\Rightarrow$ $\Pr(f_0 \to f^*) \approx 2Nsf_0 = p_{fix}(f_0)$



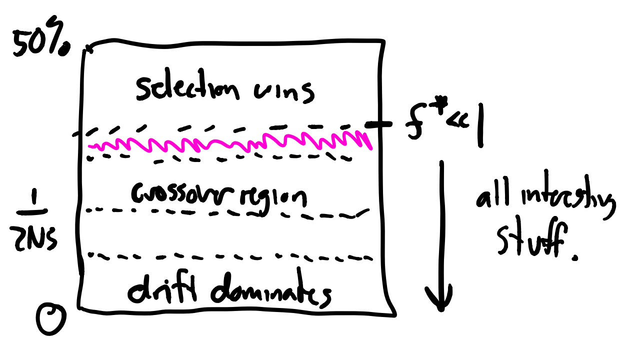
[all uncertainty in mut'ns fate takes place between $0 \le f \le f^* (\ll 1)$]

$\Rightarrow$ i.e. "selection wins" if $f(t) \gg 1/2Ns$

② if $f^* \ll 1/2Ns$ $\Rightarrow$ $p(f_0 \to f^*) = \frac{2Nsf_0}{2Nsf^*} = \frac{f_0}{f^*}$ (independent of s!)

$\Rightarrow$ i.e., similar to neutral mutation for $f(t) \ll 1/2Ns \ll 1$

$\Rightarrow$ interesting partitioning of frequency space:



$\Rightarrow$ important for evolution because new mutations: $f_0 = 1/N \ll 1/Ns$

when $f \ll 1$: evolutionary model reduces to

$$\boxed{\frac{\partial f}{\partial t} = sf + \sqrt{\frac{f}{N}}\,\eta(t) + \left[\mu - \nu f\right]}$$

known as "linear branching process"

* Intuition:

o = WT
● = mut.

All competitions occur against wildtype ($f \cdot 1$ vs $f \cdot f$)

$\Rightarrow$ Let's first consider no mutations, $\mu = \nu = 0$

$\Rightarrow$ SDE is linear (by construction)

e.g. for $\langle f(t) \rangle$ $\Rightarrow$ $\partial_t \langle f \rangle = s \langle f \rangle$ (no moment hell!)

$\Rightarrow$ $\langle f(t) \rangle = f(0)\, e^{st}$

(same as deterministic model)

$\Rightarrow$ e.g. for 1-way mutation ($\mu > 0, \nu = 0$)

$\Rightarrow$ $\partial_t \langle f \rangle = s\langle f \rangle + \mu$ $\Rightarrow$ $\langle f(t) \rangle = f(0)\, e^{st} + \frac{\mu}{s}\left(e^{st} - 1\right)$

$\Rightarrow$ e.g. $s < 0$ $\Rightarrow$ $\langle f(t) \rangle \approx \frac{\mu}{|s|}$ (just like deterministic mut-sel balance)

$\Rightarrow$ can extend to higher moments:

e.g. $\nu = 0 \Rightarrow$ can show that

$$\frac{d\langle f^2\rangle}{dt} = \underbrace{2s\langle f^2\rangle}_{\text{from deterministic part.}} + \underbrace{\overbrace{\frac{\langle f\rangle}{N}}^{\text{know this from before.}}}_{\text{from noise term.}}$$

$$\Rightarrow \langle f(t)^2\rangle = f(0)^2 e^{2st} + \frac{f(0)e^{st}(e^{st}-1)}{Ns}$$

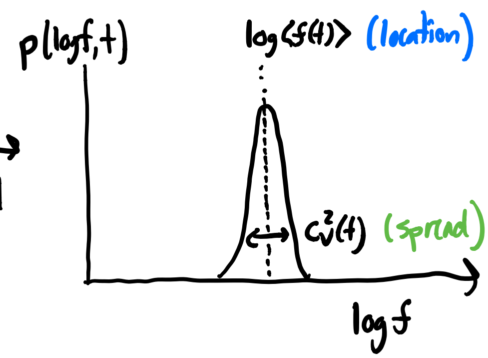
$\Rightarrow$ coefficient of variation

$$c_v^2(t) \equiv \frac{\mathrm{Var}(f(t))}{\langle f(t)\rangle^2} = \frac{\langle f^2\rangle - \langle f\rangle^2}{\langle f\rangle^2} = \frac{1-e^{-st}}{Nsf_0}$$

useful for visualizing spread of dist'n in log space

(i.e. how uncertain @ order of magnitude level?)

$\xrightarrow{\text{when } c_v(t) \lesssim 1}$

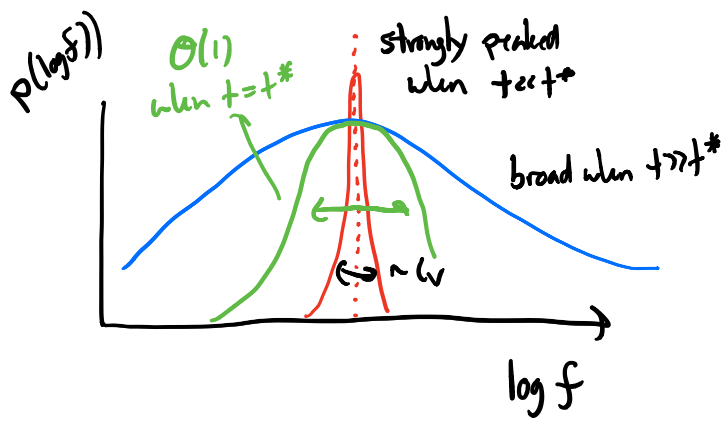


① when $s>0$ + $f_0 \gg 1/2Ns$ $\Rightarrow$ $c_v(t) \lesssim \frac{1}{Nsf_0}$

$\ll 1$ for all time!

② In contrast, when $f_0 \ll 1/Ns$, (or if $s \leq 0$)

$$C_v(t) = \frac{1-e^{-st}}{Nsf_0} \approx \begin{cases} \ll 1 & \text{if } t \ll t^* \\ \gg 1 & \text{if } t \gg t^* \end{cases} \quad \text{for some } t^*$$



can solve for t^* by setting $C_v^2(t^*) = \frac{1-e^{-st^*}}{Nsf_0} \sim 1$

can solve for t^*:

$$t^* \sim \begin{cases} \infty & \text{if } s>0,\ f_0 \gg 1/Ns \\ Nf & \text{if } f_0 \ll 1/N|s| \\ \frac{1}{s}\log(N|s|f_0) & \text{if } s<0;\ f_0 \gg \frac{1}{N|s|} \end{cases}$$

will discuss the intuition behind these expressions later!

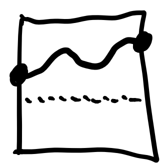
$\Rightarrow$ ideally solve for full dist'n of $f(t)$:

e.g. Fokker–Planck eq: $\frac{\partial p}{\partial t} = -\frac{\partial}{\partial f}\left[sfp\right] + \frac{1}{2}\frac{\partial^2}{\partial f^2}\left[\frac{f}{N}p\right]$

(hard!)

$\Rightarrow$ easier to focus on MGF: $H(z,t) \equiv \langle e^{-zf(t)} \rangle = \int e^{-zf} p(f,t)\,df$

$\Rightarrow$ satisfies $\frac{\partial H}{\partial t} = \left[sz - \frac{z^2}{2N}\right]\frac{\partial H}{\partial z}$ w/ $\underline{H(z,0) = e^{-zf_0}}$

solved via "method of characteristics"  (p. 4 of notes)

$\Rightarrow$ $\boxed{\text{can show that } H(z,t) = \exp\left[-\frac{f_0 z e^{st}}{1+\frac{z}{2Ns}(e^{st}-1)}\right]}$ ***

$\Rightarrow$ formally, can invert to obtain $p(f,t)$ (tricky)
(hard to interpret)

$\Rightarrow$ can learn a lot by focusing on $H(z,t)$ directly.

e.g. recall: Expand in powers of z:

$$H(z,t) \simeq 1 - z\langle f(t)\rangle + \frac{z^2}{2}\langle f(t)^2\rangle + \ldots$$

$$\Updownarrow \qquad\qquad \Updownarrow$$

$$\simeq 1 - z f_0 e^{st} \qquad \ldots$$

can rewrite generating function in suggestive form!

$$H(z,t) = \exp\left[-\frac{z\langle f(t)\rangle}{1 + \frac{z}{2}\langle f(t)\rangle \cdot c_v^2(t)}\right] \Leftarrow$$ not gaussian! $H = e^{-\mu z + \frac{1}{2}\sigma^2 z^2}$

<u>But</u>, for $t \ll t^*$ $\left[c_v^2(t) \ll 1\right]$

$$H(z,t) \simeq \exp\left[-z\langle f(t)\rangle + \frac{z^2}{2}\underbrace{\langle f(t)\rangle^2 c_v^2(t)}_{Var(f(t))}\right] \rightarrow$$ will be a Gaussian (in bulk) w/ mean $\langle f(t)\rangle$ & $c_v \ll 1$

("case 1" dist'n)

mean + spread.

outside this regime: $H(z,t) \equiv \langle e^{-z \cdot f(t)} \rangle$

"time dependent extinction prob"

when $z \to \infty$ $\quad H(z,t) \approx 0 \times (1 - p_{ext}(t)) + 1 \times p_{ext}(t)$

$$= p_{ext}(t)$$

$\Rightarrow$ in our case: $p_{ext}(t) = \exp\left[-\frac{2Nsf_0}{1-e^{-st}}\right] = \exp\left[-\frac{z}{c_v^2(t)}\right]$

can also define survival probability:

$$p_{survive}(t) \equiv 1 - p_{ext}(t) \xrightarrow[t \to \infty]{} p_{fix}(N, s, f_0)$$