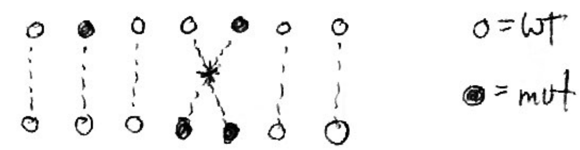
Dynamics of Linear Branching Processes

Recall when $f \ll 1$, the single locus diffusion model reduces to the linear form:

$$\frac{\partial f}{\partial t} = sf + \sqrt{\frac{f}{N}}\,\eta(t)$$

($+\mu$ for forward mutation)
($-\nu f$ for backward mutation, mut $\to$ WT)

* For intuition, helps to think about the $f \ll 1$ limit in context of the microscopic Moran model:



$\Rightarrow$ in $f \ll 1$ limit, all competitions (drift & selection) w/ mutant come @ expense of WT population (prob $f \cdot 1$)

$\Rightarrow$ competition between 2 mutant individuals is rare (prob $f \cdot f$)

hence, mutant never "feels" effects of itself in population (linearity)

* Technically, branching process model now allows f to range between 0 and ∞ $\Rightarrow$ in practice, we'll make sure to switch back to full single locus model long before $f(t)$ gets close to 50% (let alone ∞)

By construction, this SDE is now linear

So moment hell is no longer a problem.

$\Rightarrow$ e.g. for mean $\langle f(t) \rangle$, we now have

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

just like deterministic model w/ $N = \infty$

$\Rightarrow$ if add 1-way mutation ($WT \xrightarrow{\mu} mut$), then

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

$\Rightarrow$ when $s < 0$ $\quad \langle f(t) \rangle \longrightarrow \dfrac{\mu}{|s|}$

[just like deterministic mutation-selection balance in full model when $|s| \gg \mu$]

$\Rightarrow$ can extend to higher moments too. e.g. when $\mu = 0$, can show that

$$\partial_t \langle f^2 \rangle = 2s \langle f^2 \rangle + \frac{\langle f \rangle}{N}$$

solution we know from before.

$\Rightarrow$ w/ a bit of work can show that

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

which yields a compact expression for the coefficient of variation,

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

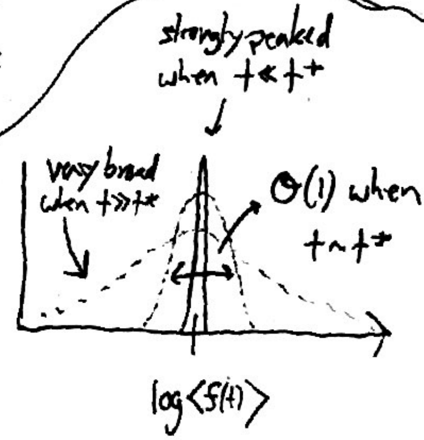
$\Rightarrow$ recapitulates results from our fixation probability discussion (now with time-dep!)

① when $s>0$ and $f_0 \gg 1/Ns$, $f(t)$ is tightly peaked around deterministic trajectory, $\langle f(t)\rangle = f_0 e^{st}$, for all time

② in contrast, when $f_0 \ll 1/Ns$, or if mutation is neutral or deleterious ($s \leq 0$) the C_v will eventually become large — and trajectory uncertain — if we wait long enough.

$\Rightarrow$ can estimate crossover time by searching for t^* where $C_v(t^*)$ first exceeds ~ 1. we find:

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



end of lecture 5

* In these cases, will want to solve for full distribution of $f(t)$

$\Rightarrow$ can in principle do this by solving time-dependent Fokker-Planck eq:

$$\frac{dp(f,t)}{dt} = -\frac{\partial}{\partial f}\left[sf\,p(f,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial f^2}\left[\frac{f}{N}p(f,t)\right] \rightarrow \text{but this is hard!}$$

(can formally solve w/ separation of variables, but infinite series / special functions hard to interpret.)

For linear branching processes, it will be much easier to focus on equivalent <u>moment generating function</u>,

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

Note: difference from full single locus model is

$$\frac{\partial H}{\partial z} - \frac{\partial^2 H}{\partial z^2} \Leftrightarrow \frac{\partial H}{\partial z}$$

$$(f(1-f)) \qquad (f)$$

which obeys the PDE: $\dfrac{\partial H}{\partial t} = \left[sz - \dfrac{z^2}{2N} \right] \dfrac{\partial H}{\partial z}$

w/ initial condition $H(z,0) = e^{-zf_0}$

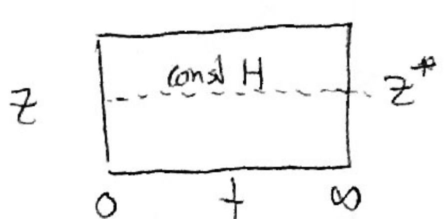
can skip to solution on page 6...

PDEs of this form can be solved w/ a technique known as the <u>method of characteristics</u>. — a generalization of the trick we used to derive the fixation probability for the full single-locus model

---

$\Rightarrow$ Recall in that case we found a special value of $z = z^*$ for which $\dfrac{\partial H(z^*,t)}{\partial t} = 0$, This allowed us to relate $H(z^*,t)$ @ sufficiently long times (where $f = 0,1$) with the initial value $H(z^*,0) = e^{-z^* f}$

if we had to draw this out in z,t plane, it would look like



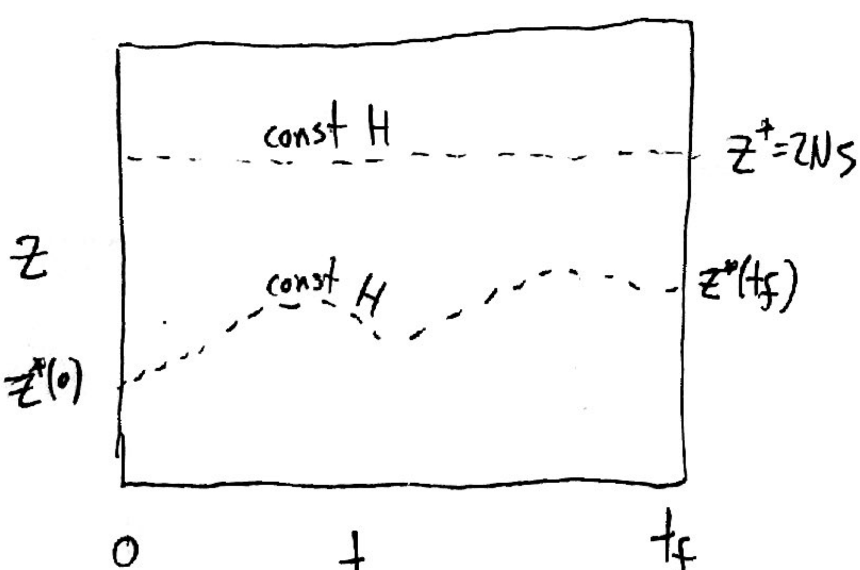
Optional: Method of characteristics

This idea can be generalized by searching for a <u>family of curves</u>, $z^*(t)$, along which
$$\frac{d}{dt}\left[H(z^*(t),t)\right] = 0$$

when this condition is satisfied, we can again relate values of $H(z,t)$ between the initial timepoint and any later time:
$$H(z^*(t),t) = H(z^*(0),0) = e^{-z^*(0) f_0}$$

The line $z^*(t) = 2Ns$ is one such <u>characteristic curve</u>, but there are infinitely many others. using the chain rule, we have
$$\frac{dH(z^*(t),t)}{dt} = \frac{\partial H}{\partial z}\frac{dz^*}{dt} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial z}\left[\frac{dz^*}{dt} + sz^* - \frac{{z^*}^2}{2N}\right]$$

so $\frac{dH}{dt} = 0$ if $\frac{dz^*}{dt} = -sz^* + \frac{{z^*}^2}{2N}$



For these curves to be any more useful than $z^* = 2Ns$, we want to be able to <u>choose</u> which value of z we evaluate H @ in the present $(t = t_f)$

e.g. $H(z_f, t_f)$

optional: method of chars...

so we need to find the value of $z^*(0)$ s.t. $z^*(t_f) = z$

$\Rightarrow$ this is easiest to accomplish by defining a function in reverse time (working back from t_f):

optional: method of characteristics

$$\varphi(t) \equiv z^*(t_f - t)$$

$\Rightarrow$ then $\varphi(t)$ satisfies $\frac{d\varphi}{dt} = s\varphi - \frac{\varphi^2}{2N}$, $\varphi(0) = z$

and $H(z,t) \equiv e^{-\varphi(t) f_0}$

$\Rightarrow$ in this case, $\varphi(t)$ is logistic equation, $\varphi(t) = \dfrac{z e^{st}}{1 + \frac{z}{2Ns}(e^{st} - 1)}$

and hence the generating function is

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

this is a key result from which much will follow

$\Rightarrow$ this generating function is tricky to invert for $p(f,t)$ [in the general case] but we can learn much by studying it directly.

Expanding in powers of z, $H(z,t) \approx 1 - z\langle f(t)\rangle + \frac{z^2}{2}\langle f(t)^2\rangle + \dots$

can recover moments we derived from SDE:

$$\langle f(t)\rangle = f_0 e^{st}, \quad C_v(t) \equiv \frac{\mathrm{Var}(f(t))}{\langle f(t)\rangle^2} = \frac{1-e^{-st}}{Nsf_0}$$

$\Rightarrow$ can rewrite $H(z,t)$ in suggestive form:

$$\boxed{H(z,t) = e^{-\frac{\langle f(t)\rangle z}{1+\frac{z}{2}\langle f(t)\rangle C_v(t)}}}$$

this distribution is <u>not</u> Gaussian. But can see that for $t \ll t^*$ $[C_v(t) \ll 1]$

$$H(z,t) \approx \exp\left[-\langle f(t)\rangle z + \frac{z^2}{2}\left[\langle f(t)\rangle^2 C_v(t)\right]\right] \longrightarrow$$

will be Gaussian in bulk w/ mean $\langle f(t)\rangle$ and $C_v \ll 1$ ("strongly peaked around deterministic value")

$\Rightarrow$ to understand what's happening outside this regime, recall that

$H(z,t) \equiv \int e^{-zf} p(f,t)\,df$ has rough interpretation as

"probability that $f \lesssim 1/z$"

in particular, if $z = \infty$, $H(z,t)$ only picks up non-zero contributions from $f(t) = 0$.

$\Rightarrow$ thus,

$$\lim_{z\to\infty} H(z,t) = \lim_{z\to\infty}\left[e^{-0\cdot z}\,\Pr(f(t)=0) + \cancel{\int_{0^+}^{\infty} e^{-zf} p(f,t)\,df}^{\,0}\right]$$

$$= P_{ext}(t) \Leftarrow \text{\underline{time dependent} extinction probability}$$

Applying to our case, we find that

$$\boxed{P_{ext}(t) = \exp\left[-\frac{2Nsf_0}{(1-e^{-st})}\right] = \exp\left[-\frac{2}{C_V(t)}\right]}$$

coefficient of variation from before.

$\Rightarrow$ can also define <u>survival probability</u> $P_{survive}(t) = 1 - P_{ext}(t)$

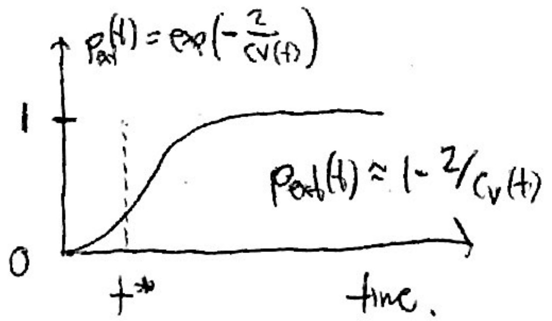
(coincides w/ fixation probability when $t=\infty$)

Our expressions have right behavior in $t\to 0$ and $t\to\infty$ limits:

1. $P_{ext}(t) \to 0$ as $t \to 0$ because we started w/ $f_0 > 0$.
2. $P_{survive}(t) \to P_{fix} = 1 - e^{-2Nsf_0}$ as $t \to \infty$ (from Kimura formula)

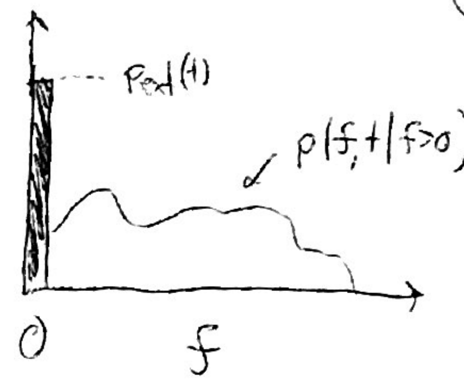
$\Rightarrow$ can now observe two classes of behavior as function of time.

1. if $s>0$ and $f_0 \gg 1/Ns$, $C_V(t) \ll 1$ and $P_{ext}(t) \approx 0$ for all time.
2. otherwise, $P_{ext}(t)$ will start ≈ 0 and will transition to $P_{ext}(t) \approx 1$ once $C_V(t) \gtrsim 1$ or $t \gg t^*$



after this point, there will be significant probability that mutant will have gone extinct: $P_{ext}(t) \approx 1 - \frac{2}{C_V(t)}$ $[t \gg t^*]$

$\Rightarrow$ can anticipate that distribution $p(f,t)$ will be "case 2" distribution that is mixture of 2 different things:



$$p(f,t) = \underbrace{P_{ext}(t)\,\delta(f)}_{\text{extinct paths}} + \underbrace{(1-P_{ext}(t))\,p(f,t\,|\,f>0)}_{\text{non-extinct paths}}$$

non-extinct paths $\rightarrow$ distribution of $f(t)$ *conditioned on survival*

↳ what we typically observe if we see mutation @ all.

$\Rightarrow$ what can we say about $p(f,t\,|\,f>0)$?

easy to calculate its mean using law of total expectation:

$$\langle f(t)\rangle = 0\cdot P_{ext}(t) + (1-P_{ext}(t))\langle f(t)\,|\,f>0\rangle$$

$$\Rightarrow \langle f(t)\,|\,f>0\rangle = \frac{f_0 e^{st}}{1-e^{-\frac{2Nsf_0}{1-e^{-st}}}} \approx \begin{cases} \langle f(t)\rangle & \text{if } t \ll t^* \\ \dfrac{e^{st}-1}{2Ns} & \text{if } t \gg t^* \end{cases}$$

$\Rightarrow$ in latter case, dependence on f_0 completely drops out! depending on selection coefficient, we obtain:

$$\boxed{\langle f(t)\,|\,f>0\rangle \xrightarrow{t\gg t^*} \begin{cases} \dfrac{1}{2Ns}e^{st} & \text{if } s>0,\ t\gg \frac{1}{s} \\ \dfrac{t}{2N} & \text{if } t \ll 1/|s| \\ \dfrac{1}{2N|s|} & \text{if } s<0,\ t\gg 1/|s| \end{cases}}$$

in other words, conditioned on survival,

1. beneficial mutations (eventually) grow exponentially w/ s, but w/ different pre-factor
2. neutral mutations grow linearly w/ time (as opposed to $\langle f \rangle = f_0$)
3. deleterious mutations eventually saturate @ const value (rather than declining exponentially)

the corresponding probabilities of these events are

$$\boxed{P_{survival}(t) \xrightarrow{t \gg t^*} \begin{cases} 2Nsf_0 & \text{if } s>0 \text{ and } t \gg 1/s \\ 2Nf_0/t & \text{if } t \ll 1/|s| \\ 2N|s|f_0 e^{-st} & \text{if } s<0 \text{ and } t \gg 1/|s| \end{cases}}$$

$\Rightarrow$ perfectly set up so that $\langle f(t) \rangle = 0 + P_{survive}(t) \langle f(t) | f>0 \rangle$

* can use similar argument to get full distribution, $p(f,t|f>0)$ via generating function:

$$H(z,t) = e^{-z \cdot 0} P_{ext}(t) + P_{survive}(t) H(z,t|f>0)$$

$$\Rightarrow \quad H(z,t|f>0) = \frac{H(z,t) - P_{ext}(t)}{P_{survive}(t)}$$

can obtain simpler expression by noting that argument of exponential in $H(z,t)$ is maximized for $z=\infty$, when $H(z,t) \approx P_{ext}(t) \approx \exp\left[-\frac{2}{C_v}\right]$

↳ thus, when $C_v \gg 1$ ($t \gg t^*$), can expand exponentials in $H(z,t)$, $P_{ext}(t)$, and $P_{survive}(t)$ to obtain:

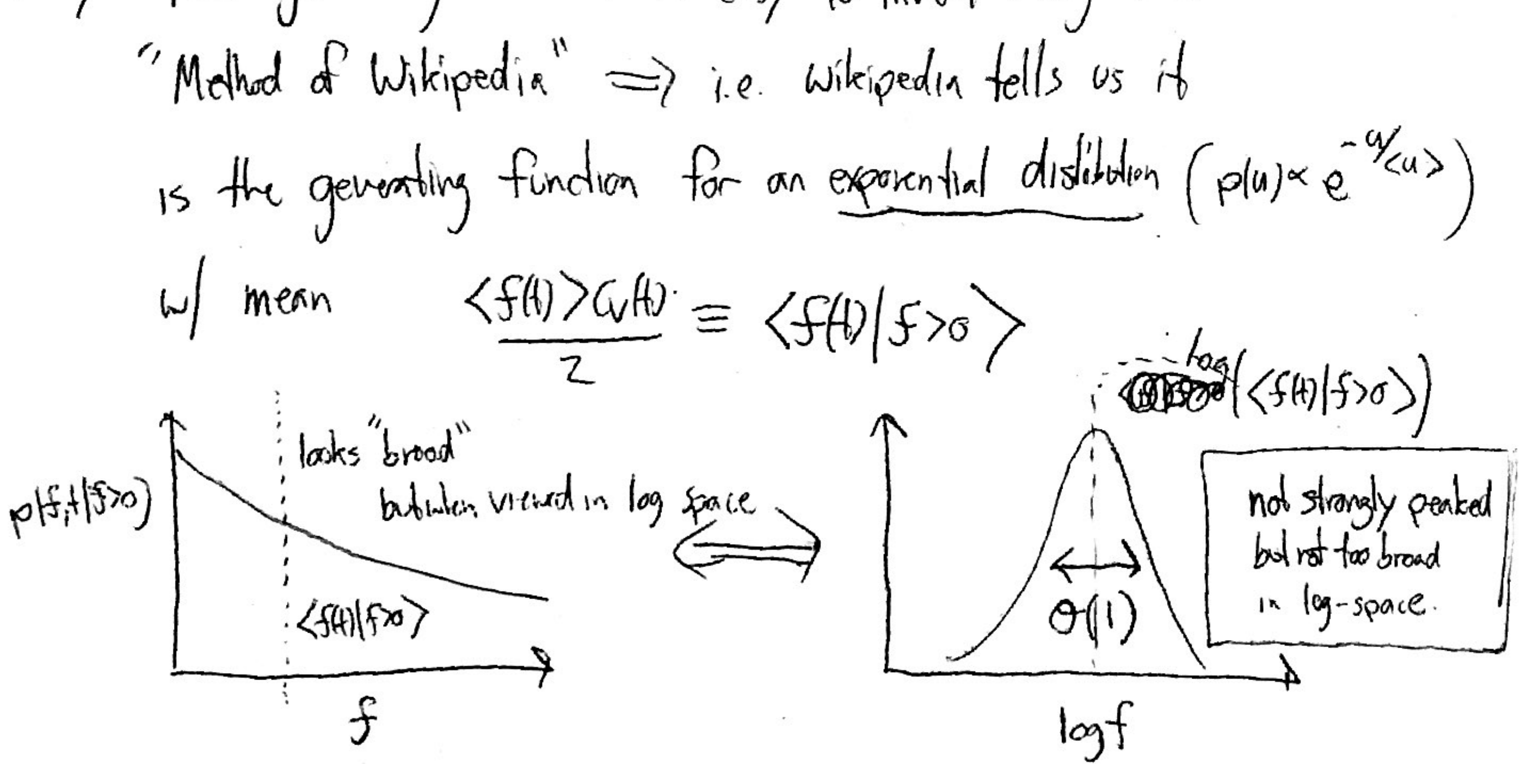
$$H(z,t) \approx 1 - \frac{z\langle f\rangle}{1+\frac{z}{2}\langle f\rangle C_v} \;;\; P_{ext}(t) \approx 1-\frac{2}{C_v} \;;\; P_{survive}(t) \approx \frac{2}{C_v}$$

Plugging into expression for $H(z,t|f>0)$, we have

$$H(z,t|f>0) \xrightarrow{t \gg t^*} \left(1+z\left[\frac{\langle f(t)\rangle C_v(t)}{2}\right]\right)^{-1}$$

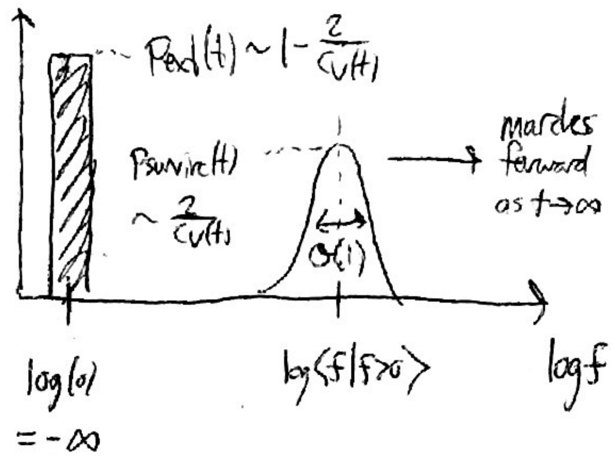
$\Rightarrow$ this generating function is easy to invert using the "Method of Wikipedia" $\Rightarrow$ i.e. wikipedia tells us it is the generating function for an exponential distribution ($p(u) \propto e^{-u/\langle u\rangle}$) w/ mean

$$\frac{\langle f(t)\rangle C_v(t)}{2} \equiv \langle f(t)|f>0\rangle$$

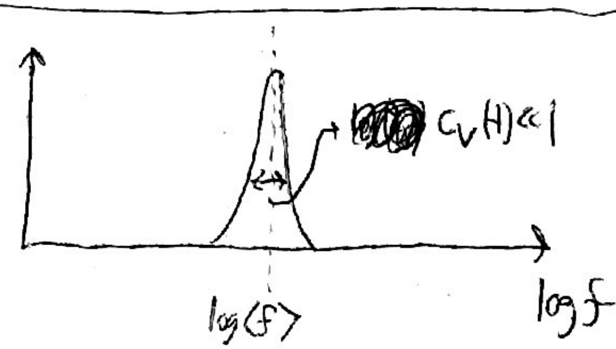


i.e. this is a distribution where the mean, $\langle f(t) | f>0 \rangle$, is a <u>reasonable</u> summary of typical behavior (i.e. up to $\mathcal{O}(1)$ prefactor)

* Returning to the full distribution, can see that for $t \gg t^*$, it breaks into the "case 2" form:



* compare to $t \ll t^*$ case:
("case 1" form)



$\Rightarrow$ can now see that for $t \gg t^*$ but $t \ll 1/|s|$ the distribution of $f(t)$ is <u>indistinguishable</u> from that of a neutral mutation, even when $N|s| \gg 1$

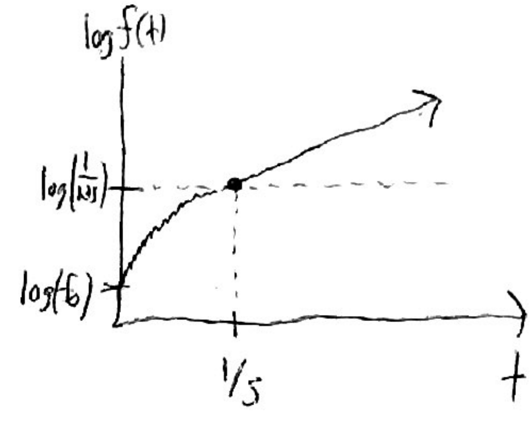
(shows that previous deduction from fixation probability extends to <u>dynamics</u> as well $\Rightarrow$ no way to tell apart!)

i.e. first

13

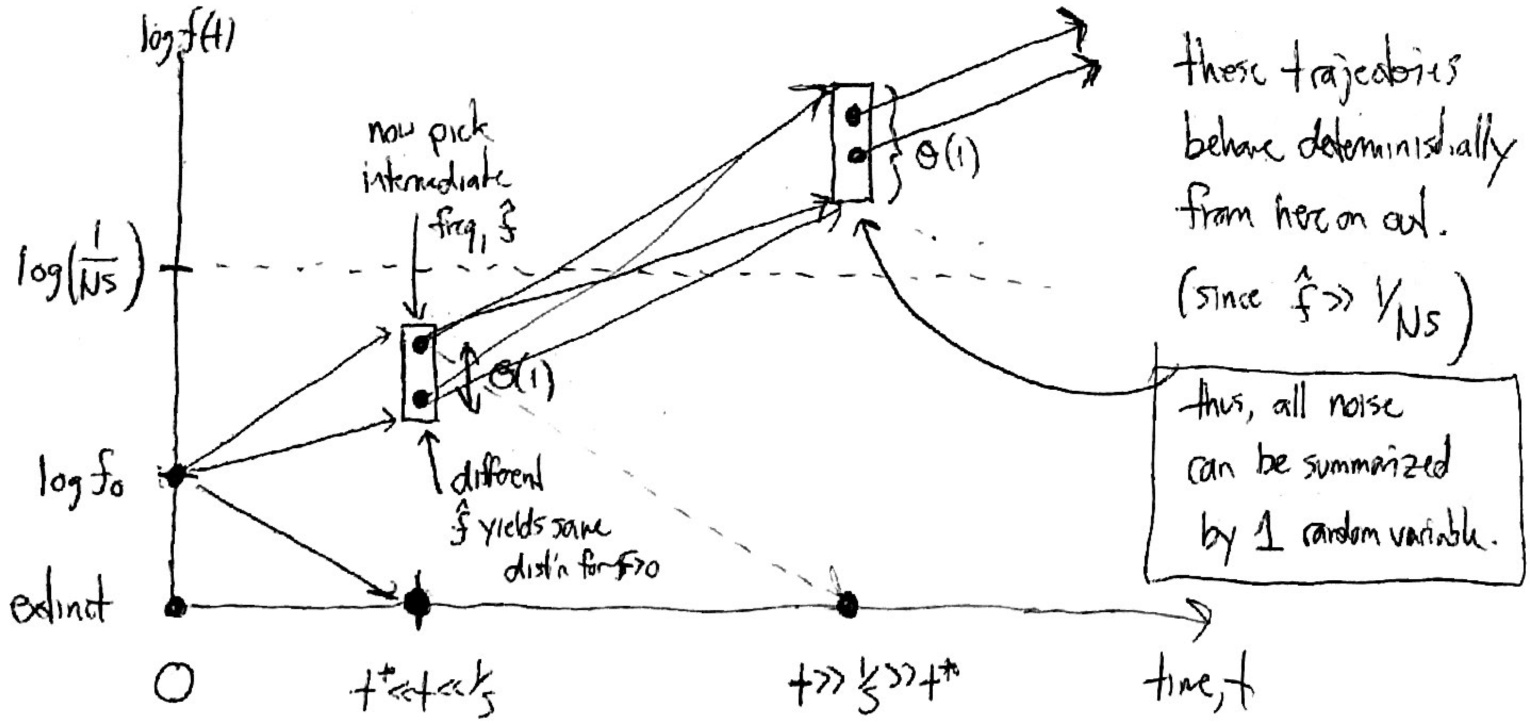
$\Rightarrow$ i.e. $t \sim Nf$ generations required to have appreciable probability of seeing $f(t) \approx f$

For beneficial mutation $(s>0)$, means that $f(t)$ grows faster than deterministically @ early times (when $t \ll 1/s$ & $f(t) \ll 1/Ns$):



$\Rightarrow$ once $t \gg 1/s$ (and $f(t) \gg \frac{1}{Ns}$), typical freqs grow deterministically w/ $\Theta(1)$ prefactor from exponential dist'n.

$\Rightarrow$ can get a better sense of what's going on by iterating and considering ensemble of paths, not just final timepoint:



Thus, for $t \gg 1/s$, all noise can be summarized by 1 random variable, $f(t) = v e^{st}$

$\hookrightarrow$ as if $f(t)$ started from different f_0

using distribution for $p(f,t|>0)$, have

$$H_v(z) = \langle e^{-zv} \rangle = \langle e^{-ze^{-st} f(t)} \rangle \xrightarrow[t \gg 1/s]{} \left(1 + \frac{z}{2Ns}\right)^{-1}$$

$\Rightarrow$ i.e. $v \sim \text{Exponential}\left(\frac{1}{2Ns}\right)$ $\left[\text{or } v \sim \frac{1}{2Ns} \cdot \underbrace{\text{Exponential}(1)}_{c}\right]$

can now do asymptotic matching to get trajectory @ later times:

Step 1 pick time t_i s.t. $t_i \gg \frac{1}{s}$ but $f(t_i) \ll \frac{1}{2}$

$\left[\text{need } \frac{1}{s} \ll t_i \ll \frac{1}{s}\log(Ns)\right]$

@ this time, $f(t_i) = v e^{st_i}$

step 2 use this timepoint as starting point for deterministic dynamics:

$$f(t) = \frac{f(t_i) e^{s(t-t_i)}}{f(t_i) e^{s(t-t_i)} + 1 - f(t_i)} \approx \frac{f(t_i) e^{s(t-t_i)}}{f(t_i) e^{s(t-t_i)} + 1} \quad \left(\text{since } f(t_i) \ll 1\right)$$

step 3 plugging in for $f(t_i)$, we have

$$\boxed{f(t) = \frac{\nu e^{st}}{\nu e^{st}+1}}$$

$\Rightarrow$ independent of t_i!
(that's how you know matching worked)

$\Rightarrow$ can use this result to calculate interesting biological quantity: time for mutation to go from $f_0 = \frac{1}{N}$ to $f(t) = 1/2$:

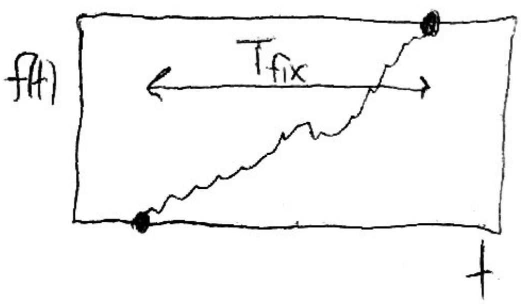
setting $f(t) = 1/2$ and solving for t, we have

$$t_{1/2} = \frac{1}{s}\log\left(\frac{1}{\nu}\right) = \frac{1}{s}\Big[\underbrace{\log(Ns)}_{\substack{\gg 1 \\ \text{(det. part)}}} + \underbrace{\log\left(\tfrac{2}{c}\right)}_{\substack{\pm\,\mathcal{O}(1) \\ \text{(random part)}}}\Big]$$

or $$\boxed{t_{1/2} \approx \frac{1}{s}\log(Ns) \pm \mathcal{O}\left(\frac{1}{s}\right)}$$

the total fixation time is just $2t_{1/2}$, or

$$\boxed{T_{fix} = \frac{2}{s}\log(Ns) \pm O\left(\frac{1}{s}\right)}$$



↳ in large populations, T_{fix} can be much larger than time it takes mutation to go from 10% → 90% (or 1% → 99%) — the "observable" part of fixation time.

⇒ will see example on homework.

↳ this quantity will be very important when we start to consider longer genomes.