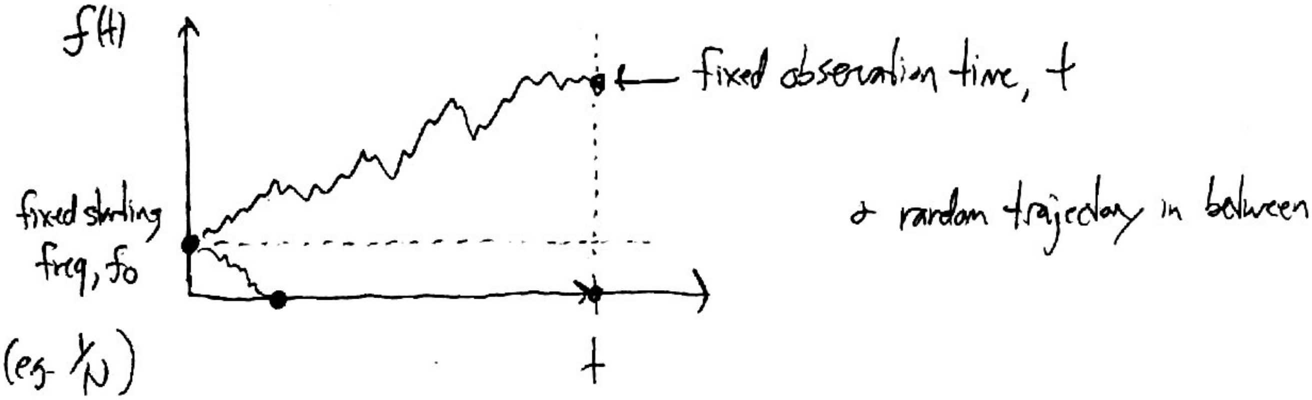
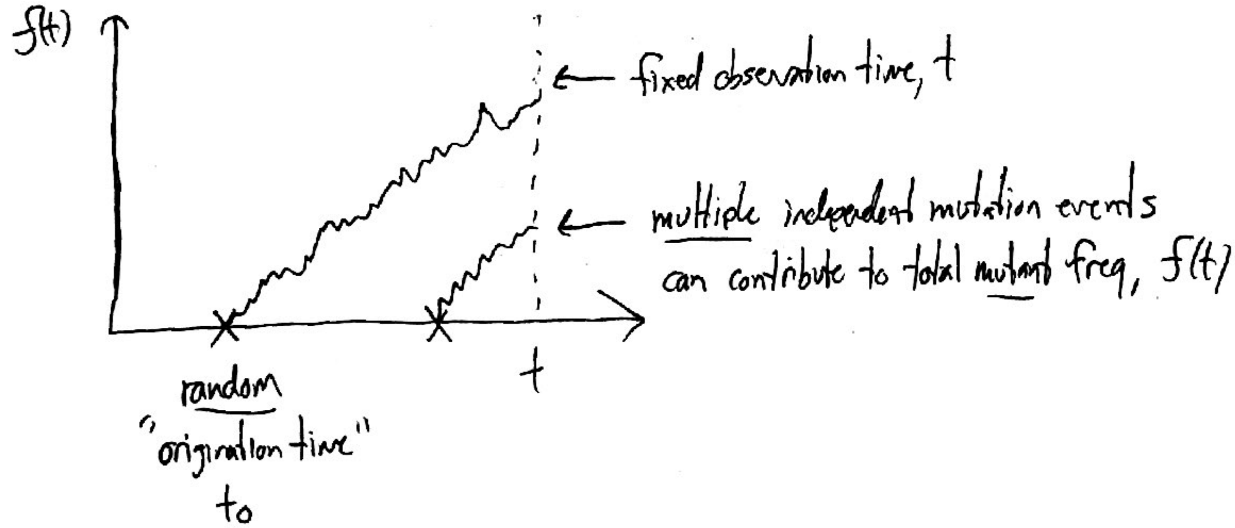
Dynamics of Linear Branching Processes II (mutations)

so far, we have focused on scenarios of the form:



in practice, often interested in frequencies of mutant types that arose spontaneously due to random mutation events, whose origination time is not known:



$\Rightarrow$ to understand this case, need to understand B.P. w/ mutations:

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

Due to linearity, can also write this process as sum over independent mutation events,

$$f(t) = \int_0^t dt_0 \sum_{i=1}^{\Theta(t_0)} f_i(t|f(t_0)=1/N) = \int_0^t dt_0\, \Theta(t_0) \times \left[\frac{1}{\Theta(t_0)} \sum_{i=1}^{\Theta(t_0)} f_i(t|f(t_0)=1/N)\right]$$

where $\Theta(t_0)$ is the (random) # of mutations produced in generation t_0 [$\Theta(t_0) \sim \text{Poisson}(N\mu)$]

and $f_i(t|f(t_0)=1/N)$ is random trajectory of <u>mutation event</u> that occurred @ generation t_0 (can predict w/ previous $\mu=0$ results)

e.g. on average,

$$\langle f(t) \rangle = \int_0^t dt_0 \underbrace{\langle \Theta(t_0) \rangle}_{N\mu} \underbrace{\langle f(t|f(t_0)=1/N) \rangle}_{\frac{1}{N} e^{s(t-t_0)}} = \frac{\mu}{s}(e^{st}-1)$$

$\Updownarrow$

same as solving $\partial_t \langle f \rangle = \mu + s\langle f \rangle$

$\Rightarrow$ what about <u>distribution</u> of $f(t)$?

$\Rightarrow$ can again return to generating function, $H(z,t) = \langle e^{-zf(t)} \rangle$

Repeating our derivation for the $\mu=0$ case, we find that $H(z,t)$ satisfies the partial differential equation,

$$\frac{\partial H}{\partial t} = \left[sz - \frac{z^2}{2N} \right] \frac{\partial H}{\partial z} - \underbrace{\mu z H}_{\text{new term.}}$$

this time, it will be most interesting to consider an initial condition with <u>no</u> mutant individuals : $H(z,0) = e^{-z \cdot 0} = 1$

this PDE can be solved using the method of characteristics (actually, a slight variant of what we did before)

can skip to solution on p. 4 below

In particular, note that if we define the function,

$$\psi(t) \equiv \log H(\underbrace{z^*(t_f - t)}_{\varphi(t)}, t_f - t),$$ where z^* is characteristic curve from $\mu=0$ case

then $\psi(t)$ satisfies the <u>ODE</u>

$$\frac{d\psi(t)}{dt} = \mu \varphi(t) \quad w/ \quad \begin{aligned} \psi(0) &= \log H(\overbrace{\varphi(0)}^{z_f}, t_f) \\ \psi(t) &= \log H(\varphi(t), 0) = 0 \end{aligned}$$

thus, we have

$$\underbrace{\psi(t)}_{0} = \underbrace{\psi(0)}_{\log H(z,t)} + \int_0^t \mu\,\varphi(t')dt' \Rightarrow \boxed{H(z,t) = e^{-\mu\int_0^t \varphi(t')dt'}}$$

plugging in for the characteristic curve, we have:

$$H(z,t) = e^{-\mu\int_0^t \frac{z e^{st'}dt'}{1+\frac{z}{2NS}(e^{st'}-1)}} = e^{-2N\mu \log\left(1+\frac{z}{2NS}(e^{st}-1)\right)\Big|_0^t} = e^{-2N\mu\log\left(1+\frac{z}{2NS}(e^{st}-1)\right)}$$

or

$$H(z,t) = \left(1+z\cdot\frac{e^{st}-1}{2NS}\right)^{-2N\mu}$$

$\Rightarrow$ can recognize as generating function for <u>Gamma distribution</u> w/ shape $\alpha = 2N\mu$ and scale $(e^{st}-1)/2NS \equiv f_{max}$

$$\boxed{p(f)\,df = \frac{1}{\Gamma(2N\mu)}\left(\frac{f}{f_{max}}\right)^{2N\mu-1} e^{-f/f_{max}}\left(\frac{df}{f_{max}}\right)}$$

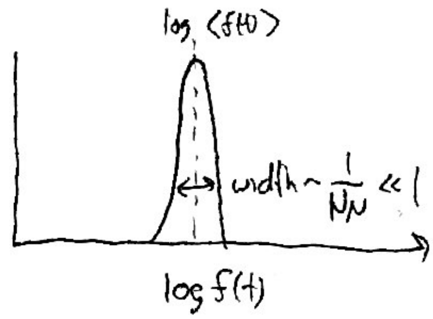
↳ <u>dynamic</u> version of mutation-selection-drift balance what does this distribution look like?

From properties of Gamma dist'n, know that

$$\langle f(t) \rangle = \alpha f_{max} = 2N\mu \cdot \left(\frac{e^{st}-1}{2Ns} \right) = \frac{\mu}{s}\left(e^{st}-1\right) \quad \left[\text{same as from SOE} \right]$$

$$\mathrm{Var}(f(t)) = \alpha f_{max}^2 = \frac{1}{2N\mu} \langle f(t) \rangle^2 \implies \boxed{C_V(t) = \frac{1}{2N\mu}}$$

$\implies$ unlike single trajectory, no longer a transition w/ t.
Instead, like equilibrium mut-sel-drift balance, 2 very diff behaviors depending on $\underline{N\mu}$



<u>case 1</u>: when $N\mu \gg 1$, dist'n strongly peaked around characteristic value, $\langle f(t) \rangle$:

e.g. for deleterious mutations $(s<0)$

$$\langle f(t) \rangle = \frac{\mu}{|s|}\left(1-e^{-|s|t}\right) \rightarrow \bar{f} = \frac{\mu}{s}$$

"deterministic mut-sel balance".

can understand from integral eq'n:

$$f(t) = \int_0^t dt_0 \; \Theta(t_0) \times \left[\frac{1}{\Theta(t_0)} \sum_{i=1}^{\Theta(t_0)} f_i\left(t \mid f_i(t_0) = 1/N\right) \right]$$

law of large #s: $\quad \langle \Theta(t_0) \rangle = N\mu \quad \times \quad \langle f_i(t \mid f_i(t_0)=1/N) \rangle = \frac{1}{N} e^{s(t-t_0)} \quad$ = deterministic value can calculate spread w/ perturbative approach:

let $f(t) = \bar{f} + \delta f(t)$ w/ $\delta f(t) \ll \bar{f}$. then plugging into SDE, obtain:

$$\frac{d(\delta f)}{dt} = -|s| \cdot \delta f(t) + \sqrt{\frac{\bar{f}}{N}}\,\eta(t) \implies$$

classic physics problem (Brownian particle in quadratic potential) w/ $r = |s|$, $D = \bar{f}/N$.

$$\implies \mathrm{Var}(\delta f) \sim \sqrt{\frac{D}{r}} \sim \frac{\bar{f}^2}{N\mu} \checkmark$$

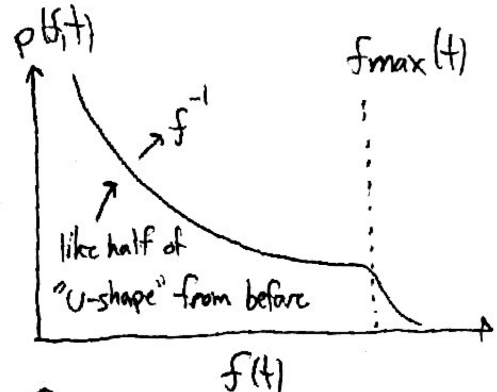
$\implies$ Note: this general approach (almost) always works when f is close to avg value. (might recognize as mean-field theory from stat phys.)

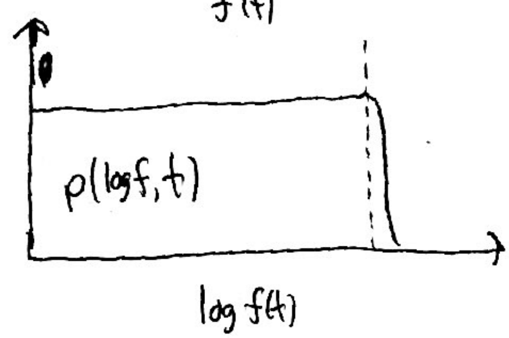
<u>case 2</u>: In opposite extreme ($N\mu \ll 1$) dist'n of f becomes very broad:

$$p(f,t) \approx 2N\mu f^{-1} e^{-f/f_{max}}$$

$\Downarrow$ or, in log space

$$p(\log f, t) \approx 2N\mu\, e^{-e^{\log(f/f_{max})}} \approx 2N\mu\, \Theta(f_{max} - f)$$





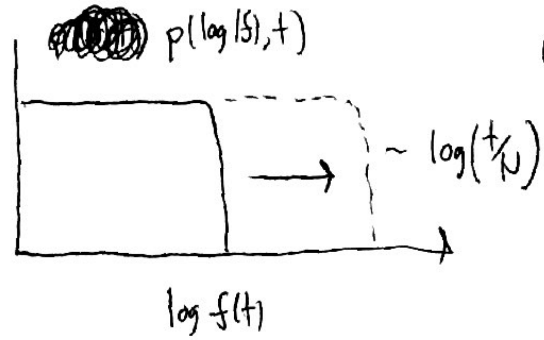
<u>Interpretation</u>: most weight near $f=0$, but w/ prob $\sim N\mu$ considerable chance of observing any order of magnitude between 0 & f_{max}

(little chance of $f \gg f_{max}$)

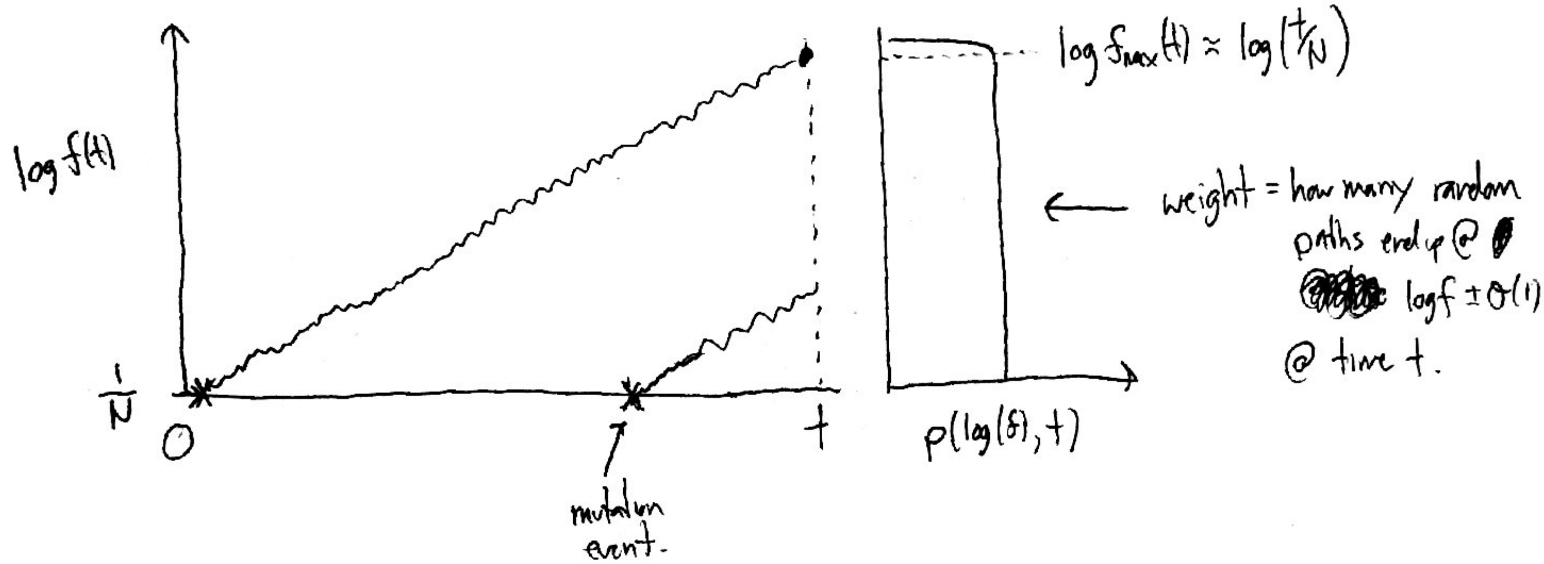
⑦

e.g. for neutral mutation, $f_{max}(t) = t/2N$

$$\Rightarrow \boxed{p(f,t) = \frac{2N\mu}{f} e^{-\frac{2Nf}{t}}}$$



* can understand dist'n as contribution from @ <u>most 1</u> random mutation event between 0 and t. Helps to visualize as:



$\Rightarrow$ can see that $f_{max} \sim t/N$ is largest size mutation could have reached if it occured @ earliest possible time ($t=0$).

$\Rightarrow$ hence, little probability of seeing $f \gg f_{max}$.

$\Rightarrow$ for $f \ll f_{max}$, we see that to contribute to $\log f \pm O(1)$, must have arisen at least $\sim Nf$ generations earlier, so that there is time to drift from $1/N \rightarrow f$.

what range of times contribute?

$\Rightarrow$ since $f \sim t/N$ $\Rightarrow$ $\log f \pm \Theta(1) \Leftrightarrow \Delta t \sim Nf$

probability of surviving until t: $\sim 1/Nf$

$\Rightarrow$ mutations w/ $t_0 \ll t - Nf$ are much less likely to be alive to contribute to $\log f \pm \Theta(1)$ @ time t.

Putting everything together, have:

$$p(\log f, t)\, \Delta \log f \simeq \underbrace{N\mu}_{\text{prob that mutation occurs per gen.}} \times \underbrace{Nf}_{\text{prob that mutation occurs in right time window to contribute to } \log(f) \pm \Theta(1)} \times \underbrace{\left(\frac{1}{Nf}\right)}_{\text{prob that mutation survives long enough to reach } \log f \pm \Theta(1) \text{ at time } t.} = N\mu \checkmark.$$

$\Rightarrow$ thus, "U"-shape in $p(f,t)$ arises because probability of surviving until $\log f(t) \simeq \log f \pm \Theta(1)$ is balanced by larger # of times that can contribute.
origin.

what about selected mutations?

$\Rightarrow$ just like single trajectory, selected mutations indistinguishable from neutral mutations when $t \ll 1/|s|$ (since $f_{max}(t) \sim t/N$)

For deleterious mutations $(s<0)$, $f_{max}(t) \to \frac{1}{2N|s|}$

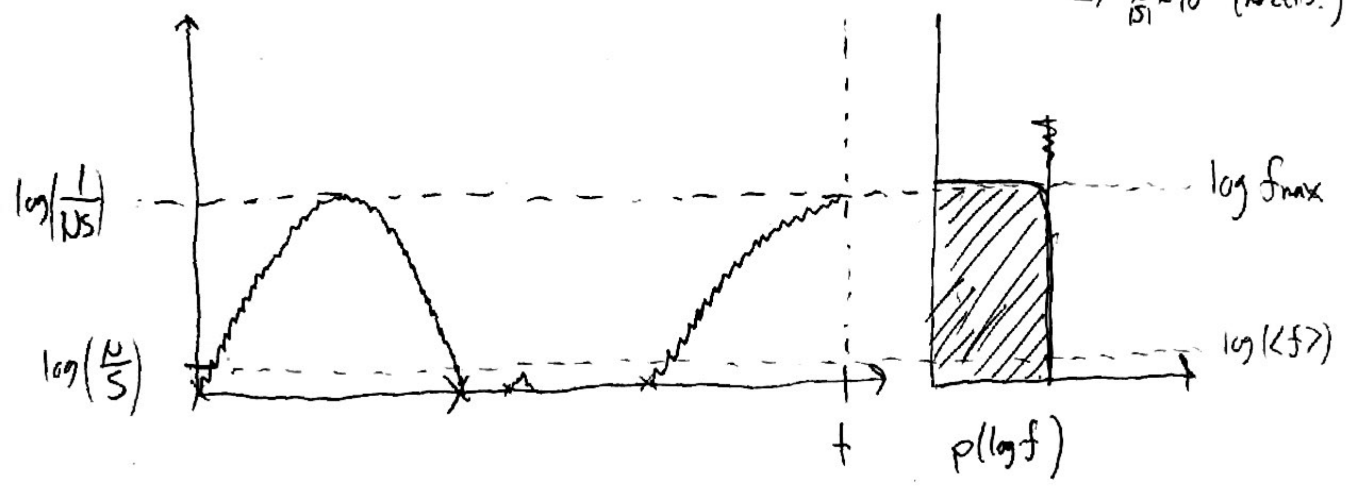
$\Rightarrow p(f,t) \to \boxed{p(f) = \frac{2N\mu}{f} e^{-2Nsf}}$ $\Rightarrow$ i.e. mostly zero, but w/ prob $\sim N\mu$ chance to grow as large as $\sim 1/N|s|$

$\Rightarrow$ very different than avg!

e.g. ABX resistance mut w/ $|s| \sim 10^{-2}$, $N \sim 10^{6}$, $\mu \sim 10^{-10}$

$\Rightarrow \frac{\mu}{|s|} \sim 10^{-8}$ (no cells!) $\frac{1}{Ns} \sim \frac{1}{10^4}$ (100 cells!)

can again visualize w/ paths:

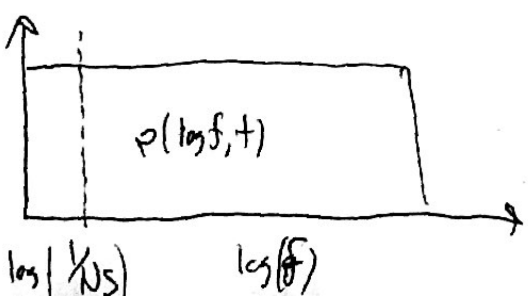


Now mutations can't get much larger than $\frac{1}{N|s|}$, can't survive much longer than $1/|s|$

$\Rightarrow$ mutations can only contribute if they arose in last $\sim 1/|s|$ gens.

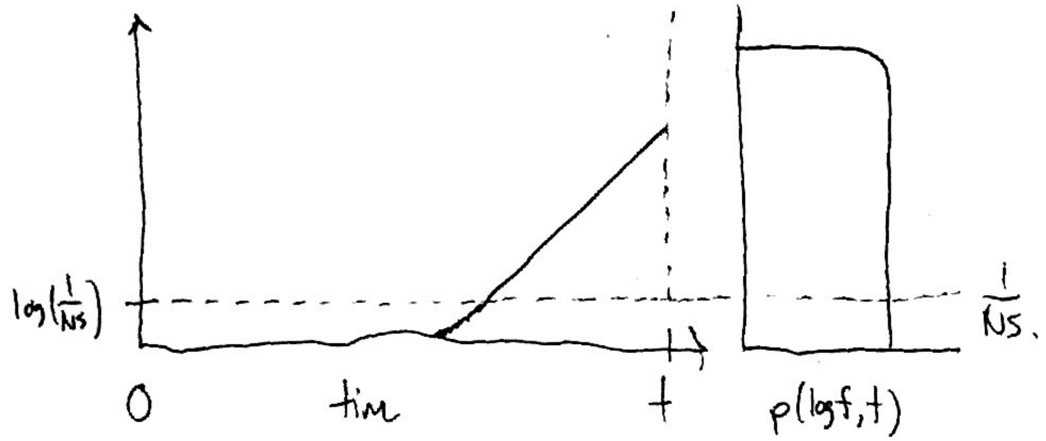
$\Rightarrow$ $\boxed{\text{hence like neutral case w/ } t_{eff} \sim 1/|s|}$

Finally, for beneficial mutations, $f_{max}(t) \to \frac{e^{st}}{2Ns} \gg \frac{1}{Ns}$



$\Rightarrow$ weird... know that $f(t)$ not drifting when $f \gg 1/Ns$.

can again understand from picture:



mutations grow as $f(t) \sim \frac{1}{Ns} e^{s(t-t_0)} \Rightarrow \log f \pm \Theta(1) \Longleftrightarrow \Delta t = \frac{1}{s}$

Putting it together: $p(\log f, t)\, \Delta \log f \simeq N\mu \times \left(\frac{1}{s}\right) \times (s) = N\mu$ ✓

$\left(\frac{1}{s}\right)$: prob that mutation occurs in right time window to contribute to $\log f \pm \Theta(1)$

(s): prob that mutation survives drift to reach $\log f \pm \Theta(1)$

$\Rightarrow$ same as neutral dist'n, but very different reason underneath.

$\Rightarrow$ differences become important when considering full path, $f(t)$.

$\Rightarrow$ path of beneficial mutation $\approx$ deterministic once $f(t) \gg \frac{1}{Ns}$.

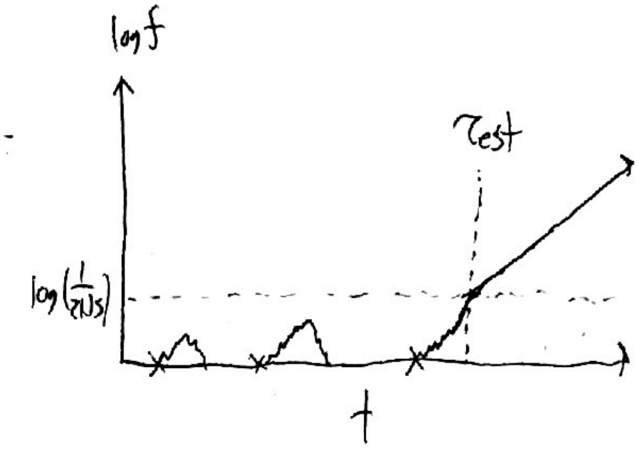
i.e. for large t, expect to capture all randomness in single #,

$$f(t) = \nu e^{st} \quad (\text{just like } \mu = 0 \text{ case})$$

$\Rightarrow$ Find $\nu \sim \text{Gamma}\left(2N\mu, \frac{1}{2Ns}\right)$ for $t \gg 1/s$. (independent of time ✓)

Often helpful to rewrite ν as a time, $f(t) = \frac{1}{2Ns} e^{s(t-\tau_{est})}$

or $\tau_{est} \equiv \frac{1}{s} \log\left(\frac{1}{2Ns\nu}\right)$



* this is known as establishment time.

$\Rightarrow$ intuitively, it's the time that $f(t)$ would have reached $\frac{1}{2Ns}$ if it grew deterministically the whole time.

(i.e. roughly time that mutation arose & survived drift)

$\Rightarrow$ can show that when $N\mu \ll 1 \Rightarrow \tau_{est} \sim \text{Exponential}\left(\frac{1}{2N\mu s}\right) \pm \mathcal{O}\left(\frac{1}{s}\right)$

Has simple interpretation: mutations occur @ rate $N\mu$ per gen & survive drift w/ prob $\sim s$.

$\Rightarrow$ successful mutations occur as Poisson process w/ rate $\sim N\mu s$.

$\Rightarrow$ "limited by supply of new mutations" (i.e. increasing N or μ by const factor decreases τ_{est} by same amount)

$\Rightarrow$ same picture also helps us understand behavior when $N\mu \gg 1$.

$\Rightarrow$ in this case, $\sim N\mu$ establishment events contribute to ν $\left(\underset{\text{rate}}{2N\mu s} \times \underset{\text{time window to contribute}}{\frac{1}{s}} = 2N\mu\right)$

each w/ typical size $\sim \frac{1}{Ns} e^{st}$.

in this case, have $\tau_{est} = -\frac{1}{s}\log(N\mu) \pm O\left(\frac{1}{s\sqrt{N\mu}}\right)$

(deterministic and negative)

$\Rightarrow$ negative because multiple mutation events contribute. initially grows much faster than e^{st}

$\Rightarrow$ time it takes mutation to reach $f = 1/2$ is $t_{1/2} \sim \frac{1}{s}\log\left(\frac{s}{\mu}\right)$

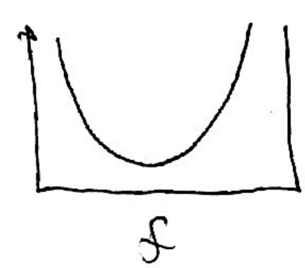
(independent of N, weakly dependent on μ $\Rightarrow$ limited mainly by strength of selection

compare to $N\mu \ll 1$ case:

$$t_{1/2} \sim \text{Exponential}\left(\frac{1}{N\mu s}\right) + \frac{1}{s}\log(Ns) \sim \frac{1}{N\mu s} \qquad \left(N\mu \ll 1/\log(Ns)\right)$$

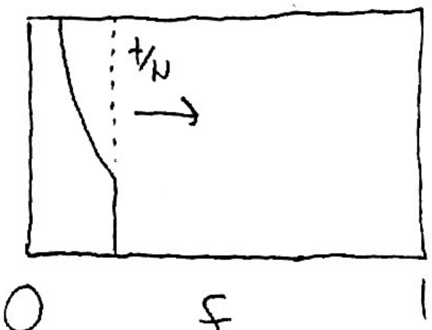
Finally, can use our new found knowledge to understand what was going on w/ "U"-shaped stationary dist'n from full single-locus model:

$$p(f) \propto f^{2N\mu-1}(1-f)^{2N\nu-1}e^{2Nsf} \xrightarrow{s=0} f^{N\mu-1}(1-f)^{N\mu-1}$$



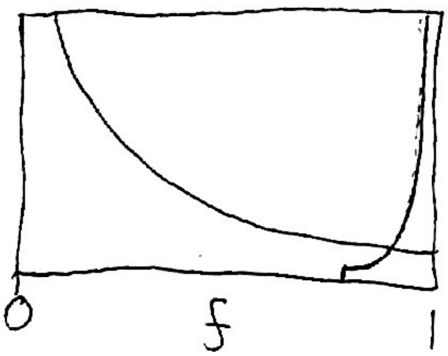
so for neutral mutations ($s=0$), movie is following:

(1)



Need $t \sim N$ generations for left half of "U-shape" to form from forward mutations from $f=0$.

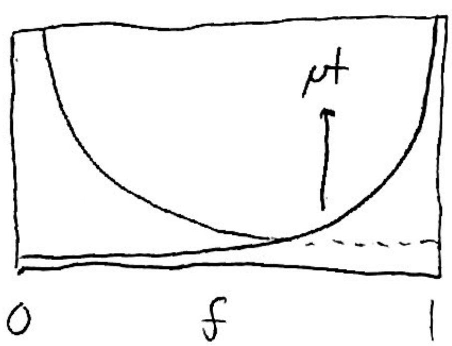
(2)



Now chance for right half of "U-shape" to form from back-mutations from $f=1$

↳ but initially, height of right half is small (since low probability to reach $f=1$)

(3)



Rate that mutations reach $f=1$ from $f=0$ is

$N\mu \times \left(\frac{1}{N}\right) = \mu \Rightarrow$ need $t_{eq} \sim 1/\mu$ generations

before $f=0$ & $f=1$ are equally likely.

* this time scale is super long! e.g. humans ~ $\mu = 10^{-8}$

so $t_{eq} \sim 10^{8}$ generations $\Rightarrow$ > 1 billion years (way longer than time since human-chimp split)

most recent common ancestor of humans lived < 10^6 years ago so not enough time for human pop to reach stationary dist'n.

Later we will see that this is true more generally:

when $N\mu \ll 1$, never enough time for neutral stationary dist'n to equilibrate in time since common ancestor of population.

$\Rightarrow$ instead, more relevant dist'n is quasi-stationary dist'n:

$$p(f) \simeq \frac{2N\mu}{f} \qquad \left(\text{valid for } s=0,\ t \gg N,\ \text{but } t \ll 1/\mu\right)$$

compare to strongly deleterious case,

$$p(f) \simeq \frac{2N\mu}{f} e^{-2N|s|f} \qquad \text{valid for } N|s| \gg 1,\ t \gg 1/|s|$$