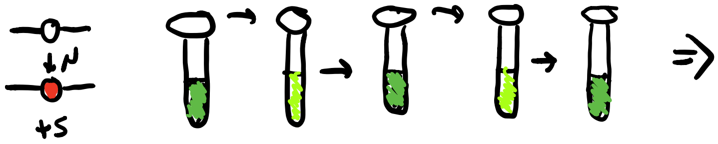
Announcements: PSET 3 Posted (DUE 2/23/21)

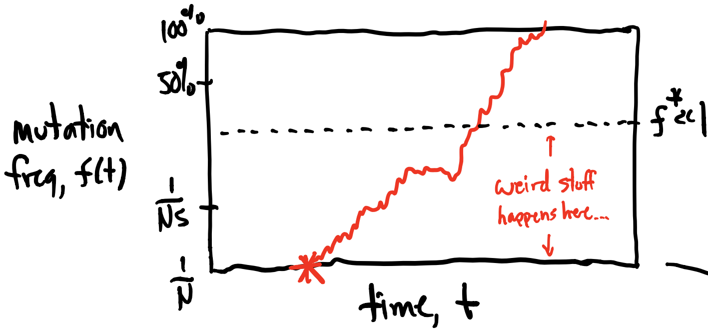
Last time: Quick review – how did we get here?



$$\Rightarrow \frac{df}{dt} = \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}$$

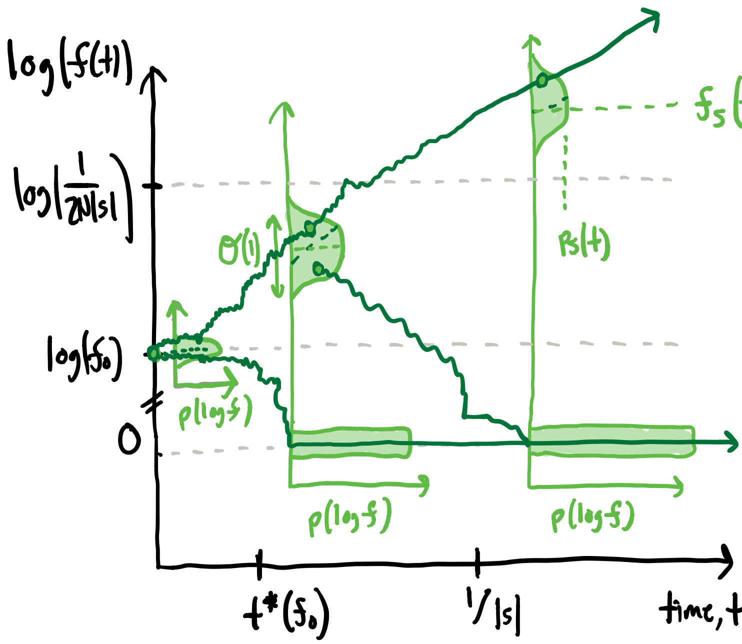
$$+\underbrace{\left[\mu(1-f) - \nu f\right]}_{\text{mutation}}$$

$$\Downarrow$$



"Linear Branching Process"

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



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

$$\approx \begin{cases} \dfrac{e^{st}}{2Ns} & \text{if } s>0;\ t \gg 1/s \\ \dfrac{t}{2N} & \text{if } t \ll 1/|s| \\ \dfrac{1}{2N|s|} & \text{if } s<0;\ t \gg 1/|s| \end{cases}$$

from generating function,

$$H(z,t) = \exp\left[-\frac{z\langle f(t)\rangle}{1+z f_s(t)}\right]$$

Today: ① Did we need all this math? ("Heuristics")

② Incorporating mutations $\xrightarrow[\text{(Thurs)}]{}$ ③ sequencing!

Heuristic approach:

$\Rightarrow$ may seem sloppy/arbitrary @ first...

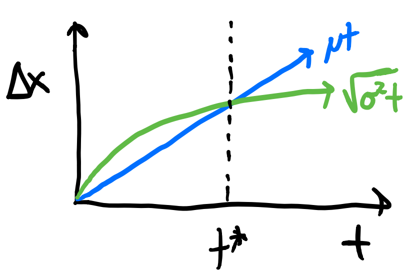
$\Rightarrow$ but w/ practice, can keep track of approx's in controlled manner while highlighting key physical intuition

$\Rightarrow$ enables progress in more complicated settings where exact results for $H(z,t)$ not available (e.g. in PSET 3)

$\Rightarrow$ start by revisiting simple Gaussian random walk

$$\frac{dx}{dt} = \mu + \sqrt{\sigma^2}\,\eta(t) \longrightarrow \text{solution } x(t) = \mu t + \sqrt{\sigma^2 t}\, \underset{}{Z}^{\sim N(0,1)}$$

$\Rightarrow$ Q: when are stochastic vs deterministic effects dominant?



$\Rightarrow$ stochastic term always dominant @ short times

$\Rightarrow$ deterministic term always dominant @ long times $(\mu \neq 0)$

$\Rightarrow$ crossover @ t^* where

$$\mu t^* = \sqrt{\sigma^2 t^*} \Rightarrow t^* = \frac{\sigma^2}{\mu^2}$$

$\nearrow$ mostly deterministic $(x \approx \mu t + \epsilon)$ when $t \gg t^*$

$\searrow$ mostly stochastic $(x \sim \sqrt{\sigma^2 t}\, Z + \epsilon)$ when $t \ll t^*$

Now return to evolution problem:

$$\frac{\partial f}{\partial t} = sf + \sqrt{\frac{f}{N}}\,\eta(t) \iff f(t+\delta t) = f(t) + \underbrace{sf(t)\,\delta t}_{\delta f_{sel}} + \underbrace{\sqrt{\frac{f(t)\,\delta t}{N}}\,Z_t}_{\delta f_{drift}}$$

$\Rightarrow$ can't apply same approach because deterministic + stochastic terms both depend on $f(t)$, which depends on det + stoch terms, etc, etc

$\Rightarrow$ needed to "integrate" SDE (moment eqs, $H(z,t)$, etc) $\Rightarrow$ HARD!!

$\Rightarrow$ Heuristics $\approx$ way to do this approximately $\approx$ "poor man's integration"

Basic idea: if interested in $\underbrace{\text{logarithmic}}_{\text{"order-of-magnitude"}}$ precision

$$\boxed{\begin{array}{c}\log f(t) \pm \mathcal{O}(1) \\ \log t \pm \mathcal{O}(1)\end{array}}$$

$\Rightarrow$ then short time approx works pretty well until $\log f(\Delta t) \approx \log(f(0)) \pm \mathcal{O}(1)$

$$\left[\, f(\Delta t) = f(0) + \underbrace{sf(0)\,\Delta t}_{\Delta f_{sel}} + \underbrace{\sqrt{\frac{f(0)\,\Delta t}{N}}\,Z_0}_{\Delta f_{drift}} \,\right]$$

$\Rightarrow$ call this time Δt_{reset}. occurs when $\log\left(\frac{f(\Delta t)}{f(0)}\right) \approx \pm\mathcal{O}(1)$ ["$\Delta f \sim f$"]

$\Rightarrow$ @ this point, set $f(0) = f(\Delta t_{reset})$ & repeat ...

$\Rightarrow$ iterative method for building up $f(t)$ for $t \gg \Delta t_{reset}$

* Question then becomes: are deterministic forces (selection) or stochastic forces (drift) dominant on timescales $\sim \Delta t_{reset}$?

$$\boxed{\Delta f = \underbrace{s f \Delta t}_{\Delta f_{sel}} + \underbrace{\sqrt{\frac{f \Delta t}{N}}\, Z}_{\Delta f_{drift}}}$$

Approach: guess & check (self consistency)

(1) if deterministic forces dominant ($\Delta f_{sel} \gg \Delta f_{drift}$)

$\Rightarrow$ then Δt_{reset} set by $\Delta f \sim f$

$\Rightarrow$ $f \sim |\Delta f| \sim |\Delta f_{sel}| \sim |s| f \, \Delta t_{reset}$

$$\Rightarrow \boxed{\Delta t_{reset} \sim T_{sel} \equiv \frac{1}{|s|}}$$

[really saying is $\Delta t_{reset} \sim \frac{c_1}{|s|}$ $\sim \mathcal{O}(1)$ const]

$\Rightarrow$ on this timescale, contribution from drift is

$$\Delta f_{drift} = \sqrt{\frac{f \Delta t_{reset}}{N}} = \sqrt{\frac{f}{N|s|}} \ll \Delta f_{sel} \sim f$$

$\Rightarrow$ good approx when $\boxed{f \gg \frac{1}{N|s|}}$ (selection dominant)

$\Rightarrow$ after 1 reset we have: $\log f(t+\Delta t) = \log f + \underbrace{\Theta(1)}_{c_2}$

$\Rightarrow$ after k resets we have:

$$\log f(t) \simeq \log f(0) + c_2 k = \log f(0) + c_3 \cdot s \cdot t$$

$\quad$ $\sim \Theta(1)$ const.

$\Rightarrow$ $f(t)$ grows exponentially @ rate $\Theta(s)$ $\left[\frac{\partial f}{\partial t} = sf \right]$

(2) if stochastic forces dominant $(\Delta f_{drift} \gg \Delta f_{sel})$

$$\Rightarrow f \sim |\Delta f| \sim |\Delta f_{drift}| \sim \sqrt{\frac{f \Delta t_{reset}}{N}}$$

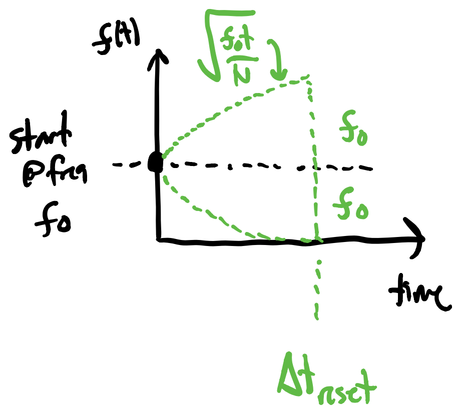
$$\Rightarrow \boxed{\Delta t_{reset} \sim T_{drift} \equiv Nf}$$

$\Rightarrow$ contribution from selection is $|\Delta f_{sel}| \sim |s| \cdot Nf \cdot f$ $(\ll \Delta f_{drift} \sim f)$

$\Rightarrow$ good approximation when $\boxed{f \ll \frac{1}{N|s|}}$ (drift dominates)

$\Rightarrow$ not simple random walk because $\sigma_{eff}^2 = \frac{f(t)}{N}$

$\Rightarrow$ but can "glue together" several ordinary random walks



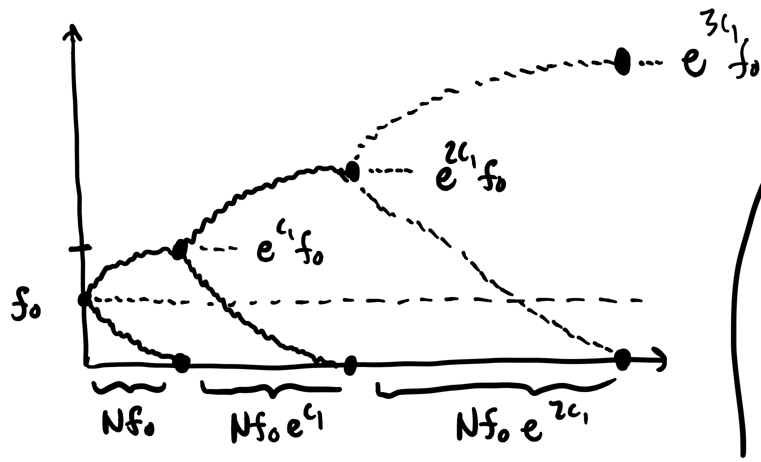
After Δt_{reset}, $f(t) \approx f_0 \pm f_0$

(decent chance of going extinct)

w/ prob $\approx e^{-c_1} \sim \mathcal{O}(1)$ factor $\approx$ [e.g. $\frac{1}{2}$]

mutation is not extinct & must have size $f \approx \frac{f_0}{e^{-c_1}} = e^{c_1} f_0$

then repeat starting from $f(0) = e^{c_1} f_0$



After k iterations

* prob of survival is $p_s \approx e^{-kc_1}$
* typical size is $f(t) \approx f_0 e^{c_1 k}$
* total time elapsed is

$t \approx Nf_0 + Nf_0 e^{c_1} + \ldots N f_0 e^{(k-1)c_1}$

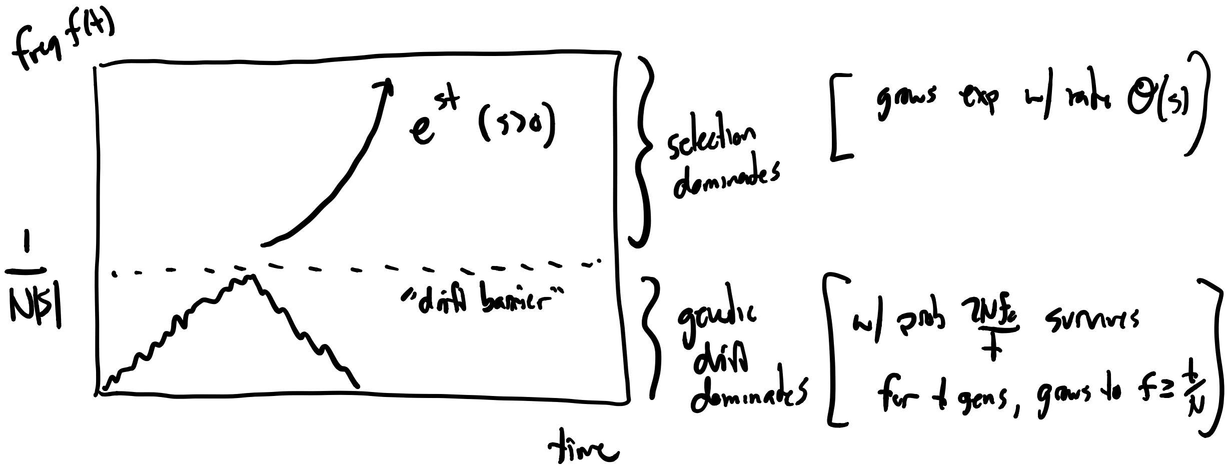
$\xrightarrow[k \gg 1]{} N f_0 e^{c_1 k}$

Rewriting in terms of time t: prob of survival is $\sim \frac{Nf_0}{t}$

typical size is $f(t) \sim \frac{t}{N}$

or in terms of final frequency $f(t) \equiv f$:

$$\boxed{\text{w/ prob} \sim f_0/f\text{, drifts to size} \sim f \text{ on timescale } t \sim Nf}$$



$\Rightarrow$ Heuristic approach assumes that boundary is sharp ("patching")

① For beneficial mutations: drifts to size $\sim \frac{1}{Ns}$

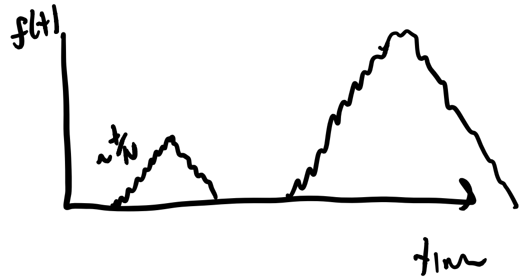
with probability $\sim \frac{1}{N} / \frac{1}{Ns} \sim s$

$\Rightarrow$ takes $\sim \frac{1}{s}$ gens to do so $\Rightarrow$ then grows exp @ rate $\Theta(s)$

② For deleterious mutations: drifts to $\sim \frac{1}{N|s|}$ w/ prob $\sim |s|$

$\Rightarrow$ prob of survival is $p_s(t) \sim |s| e^{-c \cdot s \cdot t} \to 0$

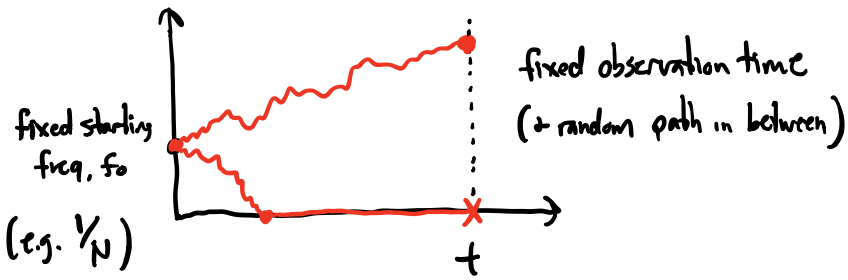
③ Neutral mutations



$\Rightarrow$ "triangles" w/ height t/N, width t, prob $\sim 1/t^2$

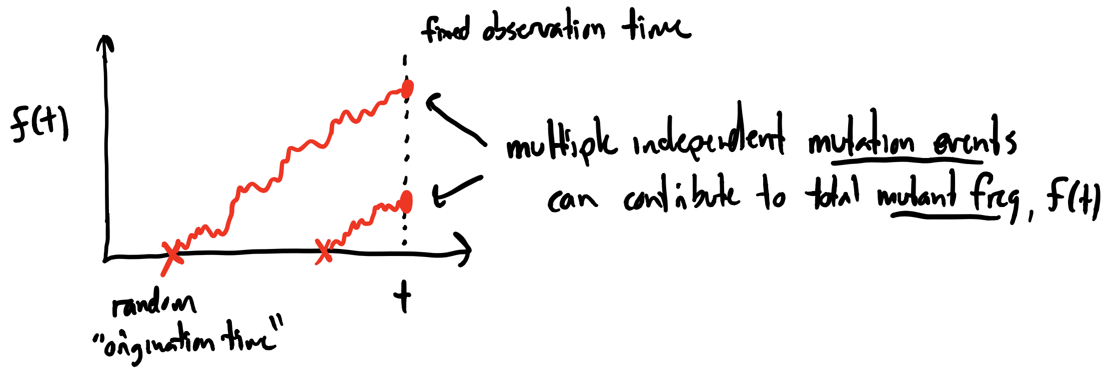
Incorporating spontaneous mutations

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



$$\boxed{\frac{df}{dt} = \underline{sf} + \sqrt{\frac{f}{N}}\,\underline{\eta(t)}}$$

⇒ in practice, often more interested in scenarios like:



LBP w/ mutation:

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

- or -

$$f(t) = \int_0^t dt_0 \sum_{i=0}^{\mathcal{B}(t_0)} \underbrace{f_i\left(t \mid f(t_0) = 1/N\right)}$$

$\mathcal{B}(t_0) \sim \text{Poisson}(N\mu)$

$$\frac{df_i}{dt} = sf_i + \sqrt{\frac{f_i}{N}}\,\eta(t)$$

⇒ can solve w/ method of characteristics $\left(H(z,t)\right)$ p. 3 of notes

$\Rightarrow$ solution $f(t)$ is Gamma distribution w/ shape $\alpha = 2N\mu$

$$f_{max} \equiv f_s(t) = \frac{e^{st}-1}{2Ns}$$

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

↳ ∗ dynamic version of mutation-selection-drift balance ∗

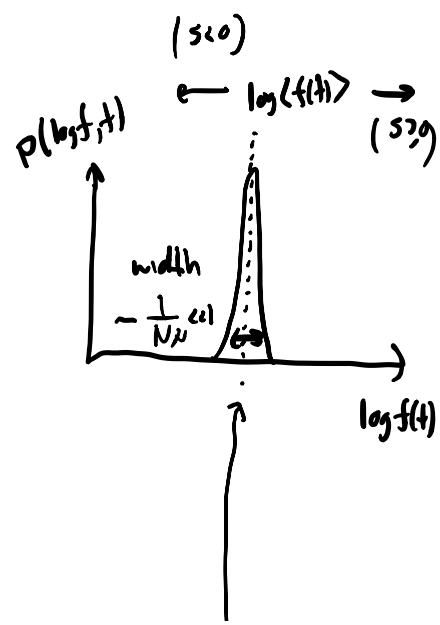
what does this look like?

$\Rightarrow$ from Wikipedia: $\langle f(t)\rangle = \alpha f_{max} = \frac{\mu}{s}(e^{st}-1)$

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

$$\Rightarrow \boxed{C_V^2(t) = \frac{Var(f)}{\langle f\rangle^2} = \frac{1}{2N\mu}}$$

case 1 when $N\mu \gg 1$, dist'n is strongly peaked around $\langle f(t)\rangle$



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

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

deterministic mut-sel balance.

can understand from:

$$f(t) \approx \int_0^t dt_0 \underbrace{\Theta(t_0)}_{\approx N\mathcal{N}} \times \left[\underbrace{\frac{1}{\Theta(t_0)} \sum_{i=1}^{\Theta(t_0)} f_i(t \mid f(t_0)=\tfrac{1}{N})}_{\approx \langle f_i(t)\rangle = \frac{1}{N} e^{s(t-t_0)}}\right]$$

law of large #s:

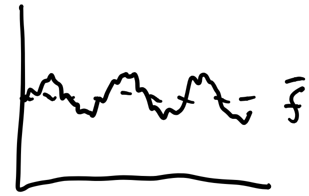
$\Longrightarrow$ can calculate spread w/ perturbative approach:

Let $f(t) = \bar{f} + \delta f(t) \Longrightarrow$ plug into SDE:

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

$\Longrightarrow$ Taylor expand in $\delta f(t) \ll \bar{f}$:

$$\Rightarrow \quad \frac{\partial\, \delta f}{\partial t} = -|s|\,\delta f + \sqrt{\frac{\bar{f}}{N}}\,\eta(t)$$



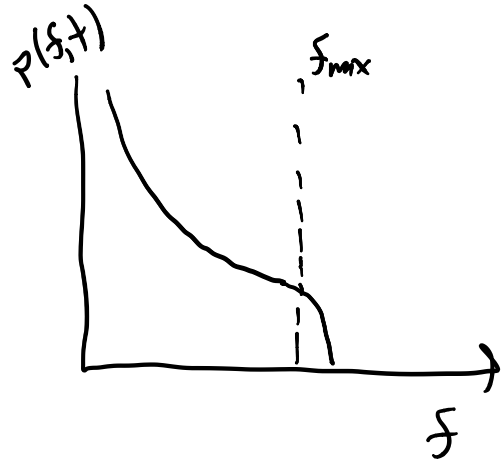
$\Rightarrow$ same as classic "Brownian particle in quadratic potential"

w/ $\bar{x} = 0$, $r = |s|$, $D = \bar{f}/N$

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

$\boxed{\text{case 2}}$ when $N\mu \ll 1 \Rightarrow$ dist'n of is very broad

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



$\Leftrightarrow$