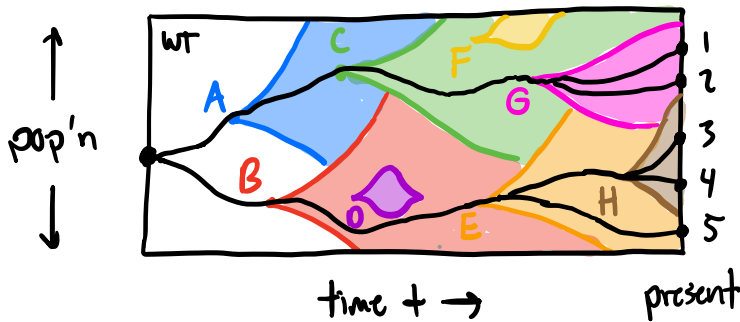


Chapter 15

Linked selection and clonal interference

Linked selection + clonal interference

(a.k.a. "Hill-Robertson Interference")



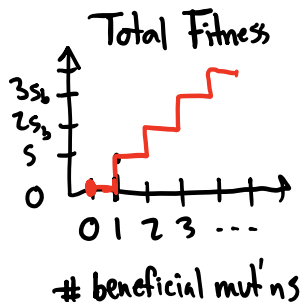
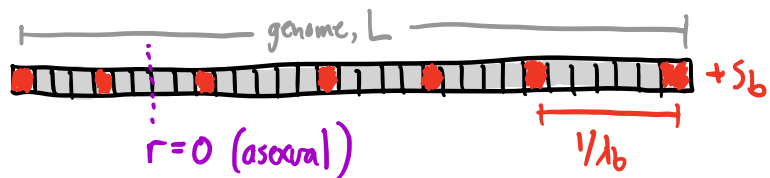
⇒ can't be reduced to $L=1$ or $L=2$ model (collective phase)

⇒ Most progress only recently, w/ big contribution from physicists

[e.g. Tsimring et al PRL '96, Rouzine et al '03, Desai + Fisher '07, ...]

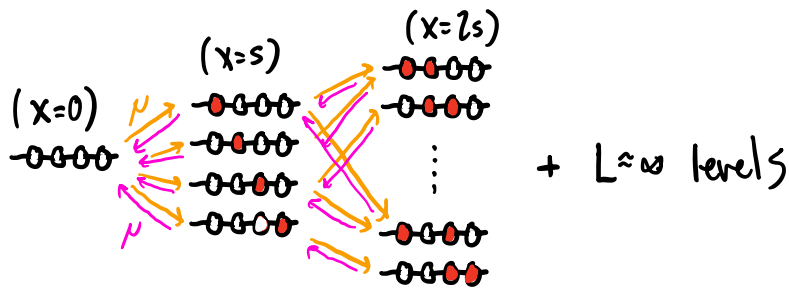
⇒ Analytical progress enabled by starting w/ very simple model:

"Staircase" Model

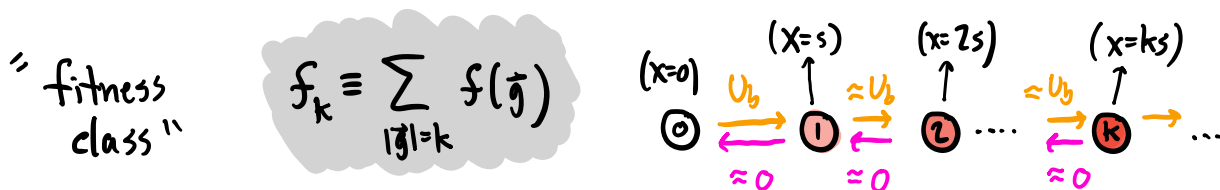


- ① All mutations provide same benefit (S_b)
- ② Occur @ total rate $U_b \equiv L\lambda_b\mu$
- ③ Never run out (e.g. $L\lambda_b \rightarrow \infty, \mu \rightarrow 0$)

Genotype network:



Key simplification:



\Rightarrow coarse-grained SDE (1+1 dimensional vs 2^L+1 dim.)

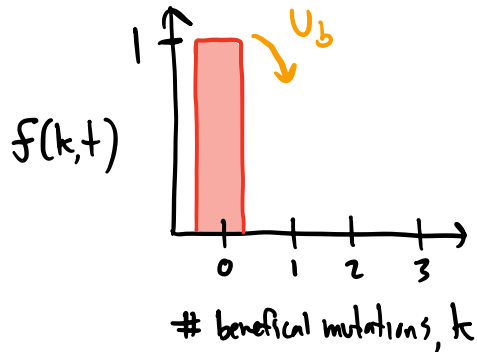
$$\frac{df(k)}{dt} = \underbrace{s_b(k - \bar{k}(t)) f(k)}_{\text{selection (nonlinear)}} + \underbrace{U_b [f(k-1) - f(k)]}_{\text{mutation}}$$

$$+ \underbrace{\sqrt{\frac{f(k)}{N}} \eta(k) - f(k) \sum_{k'} \sqrt{\frac{f(k')}{N}} \eta(k')}_{\text{genetic drift (stochastic)}}$$

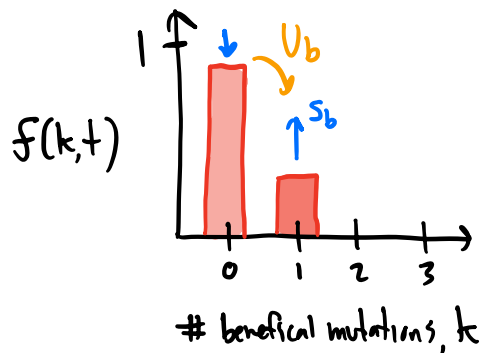
\Rightarrow let's consider behavior when $Ns_b \gg NU_b \gg 1$

(e.g. yeast barcode experiment in HW 4 Problem #1)

① Start w/ wildtype population @ $t=0$



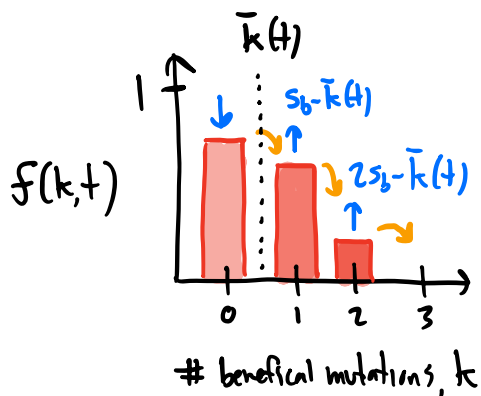
② First-step mutations ($k=1$) establish & grow exponentially



$$\Rightarrow f(1,t) \approx \frac{U_b}{s_b} (e^{s_b t} - 1)$$

(deterministic approx good @ first, since $NU_b \gg 1$)

③ Double mutants establish before single mutants take over,




\Rightarrow clonal interference!

(since $\int_0^{z_{1/2}} Ns(1,t) \cdot U_b \cdot s_b dt \sim NU_b \gg 1$)

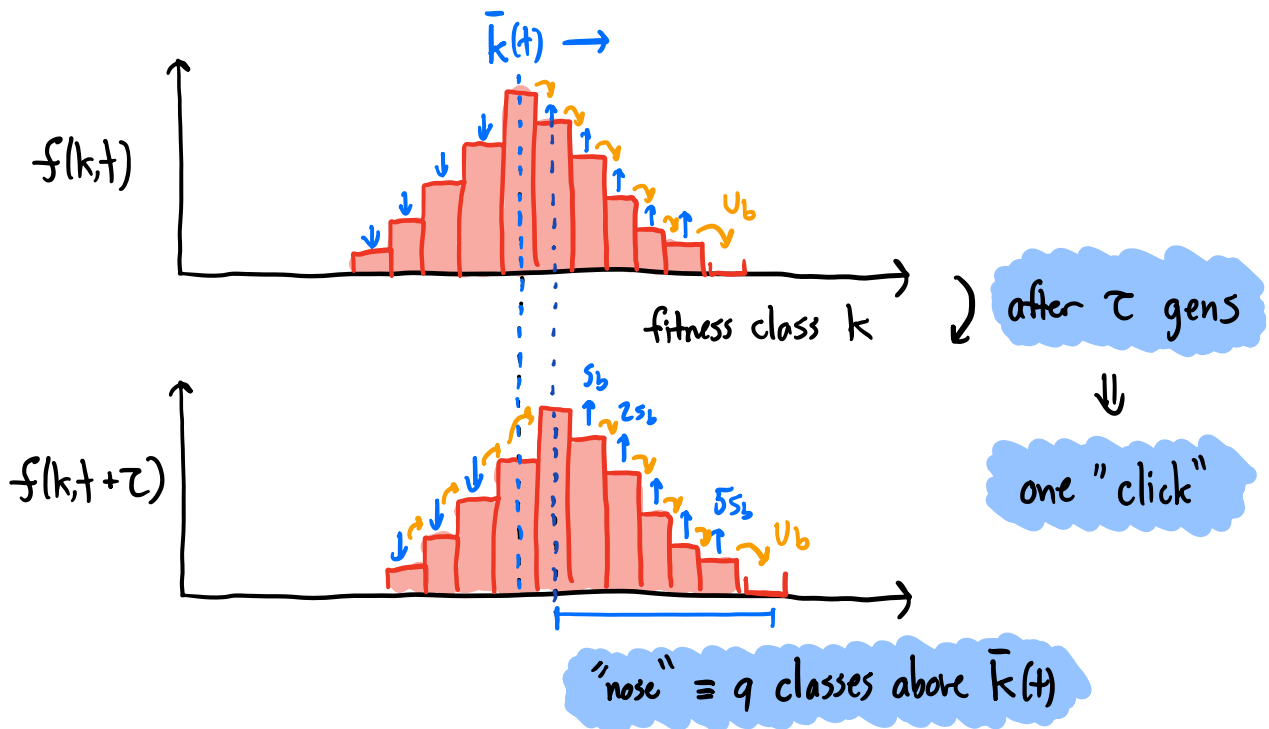
⇒ Is deterministic approx still useful?

$$\frac{df(k)}{dt} = \underbrace{s_b(k - \bar{k}(t))f(k)}_{\text{selection (nonlinear)}} + \underbrace{U_b[f(k-1) - f(k)]}_{\text{mutation}} + \underbrace{\left[\sqrt{\frac{f(k)}{N}}\eta(k) - f(k) \sum_k \sqrt{\frac{f(k)}{N}}\eta(k) \right]}_{\text{genetic drift}}$$

⇒ can show: $f_{\text{det}}(k,t) = \frac{1}{k!} \cdot \left[\frac{U_b}{s_b} (e^{s_b t} - 1) \right]^k \cdot e^{-\frac{U_b}{s_b} (e^{s_b t} - 1)}$ 

⇒ **Not self-consistent!** ⇒ Predicts $s_b \bar{k}(t) \approx U_b e^{s_b t}$
 (eventually all $f(k,t) \ll 1/N!$)

⇒ Instead, if we simulate model, observe "travelling wave":



⇒ What determines $\tau(N, s_b, U_b) + q(N, U_b, s_b)$?

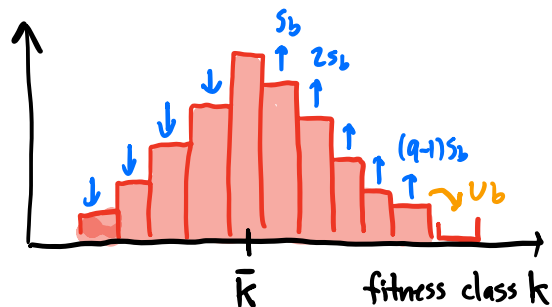
⇒ Today: heuristic analysis [\sim Desai + Fisher 2007]

applies when: $Ns_b \gg NU_b \gg 1$ + $s_b \tau \gg 1$ & $q \gg 1$

Leads to simplifications:

① mutations only important for establishing new "nose"

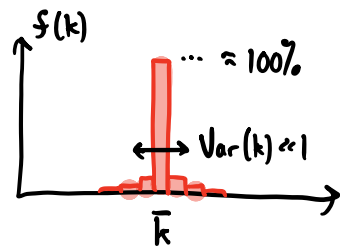
(since $s_b \gg U_b$)



② Genetic drift only important for establishing new nose

(since $\tau \gg 1/s_b$, individual mutations establish before next click.)

③ most of pop'n is near $k \approx \bar{k}(t)$

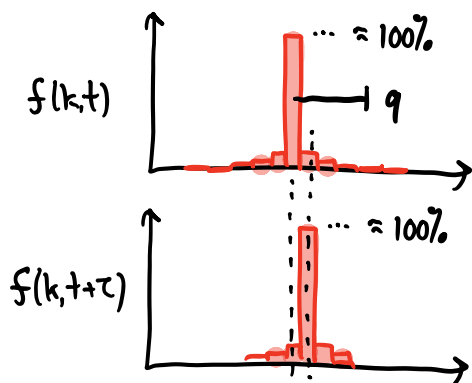


Problem 3 of HW 4:

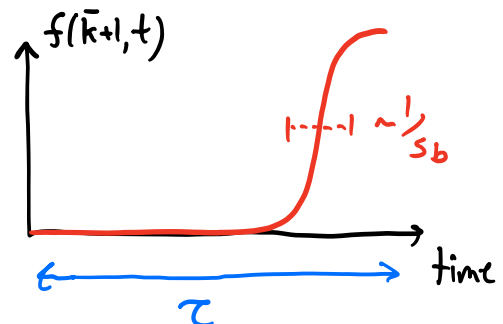
$$\underbrace{\frac{d\langle \bar{k} \rangle}{dt}}_{\equiv 1/\tau} = \left\langle \sum_k k \frac{df(k)}{dt} \right\rangle = \left\langle \underbrace{\sum_k s_b (k - \bar{k})^2 f(k, t)}_{s_b \text{Var}(k)} \right\rangle$$

$$\Rightarrow \text{Var}(k) = \frac{1}{s_b \tau} \ll 1 \quad (\text{by assumption})$$

④ Also implies that $\bar{k}(t)$ clicks suddenly:



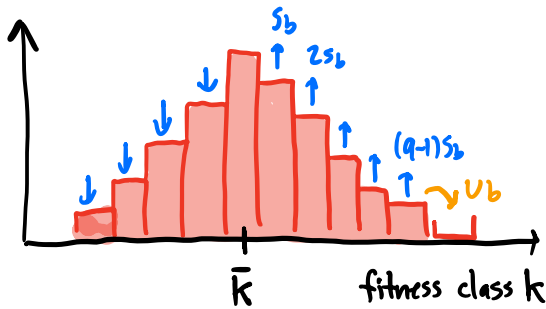
\Rightarrow



\Rightarrow i.e. for most $t \in [0, \tau] \Rightarrow \bar{k}(t) = \bar{k}(0)$

\Rightarrow everyone grows as $f(k, t) \sim f(k, 0) e^{(k - \bar{k}(0))st}$

\Rightarrow Now we have all ingredients to understand wave:

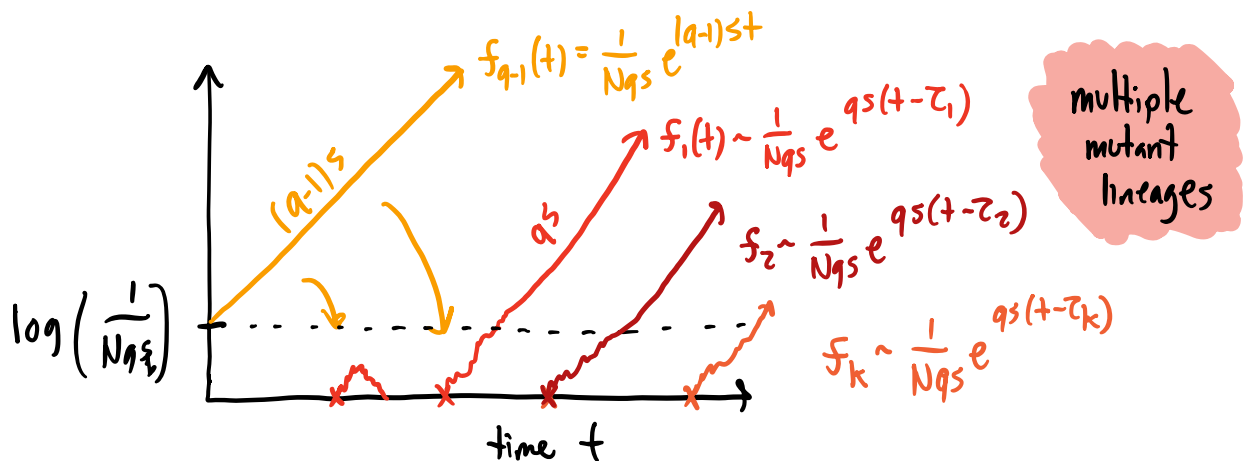


\Rightarrow in one click (τ), must establish new nose

$$\Downarrow$$

$$f_q(t) = \frac{1}{Nq_s} e^{qs(t-\tau)}$$

$\Rightarrow \tau \approx$ establishment time of nose class!



$$\Rightarrow f_q(t) = \sum_{k=0}^{K_{\max}} f_k(t) \equiv \frac{1}{Nq_s} e^{qs(t-\tau)}$$

\downarrow
 establishment time
 for whole class.

$\Rightarrow k^{\text{th}}$ successful mutant establishes when:

$$\int_0^{\tau_k} NU_b \cdot f_{q-1}(t) \cdot qs_b dt \sim \mathcal{O}(k)$$

Note: extra
 little bit
 will be
 important
 below!

$$\int_0^{\tau_k} NU_b \cdot \frac{1}{Nq_s} e^{(q-1)s_b t} \cdot qs_b dt = \frac{U_b}{qs_b} e^{(q-1)s_b \tau_k} \sim \mathcal{O}(k)$$

$$\Rightarrow \tau_k = \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b} \cdot q \cdot k\right)$$

\Rightarrow Note: $\tau_k = \underbrace{\frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b} \cdot q\right)}_{\tau_1} + \underbrace{\frac{1}{(q-1)s_b} \log(k)}_{\tau_k - \tau_1}$

(most time spent waiting for first mut'n)

\Rightarrow many mutations establish in quick succession ($\delta t \sim \frac{1}{q s_b} \ll \tau$)

\Rightarrow Typical size of k^{th} lineage:

$$\Rightarrow f_k(t) = \frac{1}{N q s} e^{q s (t - \tau_k)} = \frac{e^{q s t}}{N q s} \left(\frac{s_b q k}{U_b} \right)^{-1 - \frac{1}{q}}$$

important!

\Rightarrow Size of entire nose class:

$$\Rightarrow f_q(t) = \sum_{k=1}^{K_{\text{nose}}} f_k(t) = \frac{1}{N q s} e^{q s t} \left(\frac{s_b}{U_b} \right)^{-1 - \frac{1}{q}} \sum_{k=1}^{K_{\text{nose}}} \frac{1}{q} \frac{1}{k^{1 + \frac{1}{q}}}$$

$$\equiv \frac{1}{N q s} e^{q s (t - \tau)}$$

↙ set equal!

$$\Rightarrow \tau = \frac{(1 + \frac{1}{q})}{q s_b} \log \left(\frac{s_b}{U_b} \right) \approx \frac{1}{q s_b} \log \left(\frac{s_b}{U_b} \right)$$

vs $\tau_k \equiv \frac{1}{q s_b} \log \left(\frac{s_b}{U_b} \cdot q \cdot k \right)$] $\tau < \tau_k$ because multiple mutations contribute @ once

⇒ how to determine q ?

⇒ follow new nose over time:

$$f(\bar{k}+q, \tau) \approx \frac{1}{Nq_s} \xrightarrow{\tau} \frac{1}{Nq_s} e^{(q-1)s\tau} \xrightarrow{\tau} \frac{1}{Nq_s} e^{(q-1)s\tau + (q-2)s\tau} \rightarrow \dots$$

(right after est.)

⇒ After q clicks, old nose is new mean! (majority of pop'n)

$$f(q\tau) \sim \frac{1}{Nq_s_b} e^{(q-1)s\tau + (q-2)s\tau + \dots + s\tau} \sim \frac{1}{Nq_s_b} e^{\frac{q^2 s \tau}{2}} \sim \mathcal{O}(1)$$

⇒ system of 2 eqs for τ & q :

$$\frac{q^2 s_b \tau}{2} \approx \log(Ns_b) + \tau = \frac{1}{q s_b} \log\left(\frac{s_b}{v_b}\right)$$

$$\Rightarrow \text{solution: } q = \frac{2 \log(Ns_b)}{\log\left(\frac{s_b}{u_b}\right)} ; \tau = \frac{1}{2s_b} \frac{\log^2\left(\frac{s_b}{u_b}\right)}{\log(Ns_b)}$$

$$\Rightarrow \left\langle \frac{d\bar{x}}{dt} \right\rangle = \frac{s_b}{\tau} = \frac{2s_b^2 \log(Ns_b)}{\log^2\left(\frac{s_b}{u_b}\right)}$$

(compare to $\sim Nu_b s_b^2$ in successive mutations regime)

\Rightarrow Self consistency: $s\tau \gg 1$ & $q \gg 1$

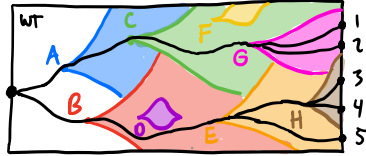
$$\Rightarrow \log\left(\frac{s_b}{u_b}\right) \ll \log(Ns_b) \ll \log^2\left(\frac{s_b}{u_b}\right)$$

Note: used heuristic derivation here...

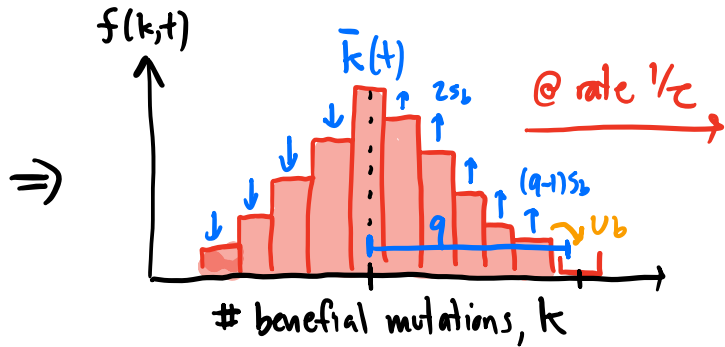
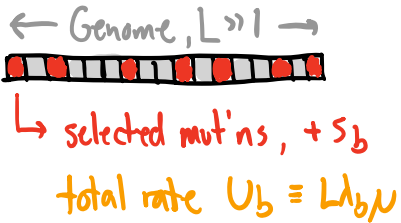
for formal analysis (using branching processes)

see Appendix A and B below

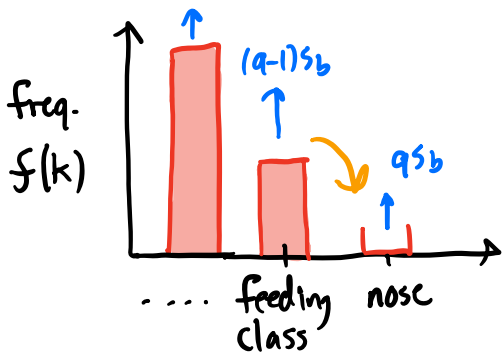
Recap: clonal interference



"Staircase" model

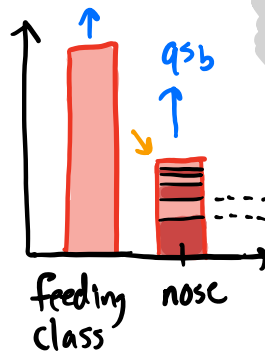


key behavior occurs @ "nose":



$+ \tau$

\Rightarrow



Multiple mutations contribute to nose!

$$f_k(t) \sim \frac{1}{Nq^s} e^{qs(t-\tau_k)}$$

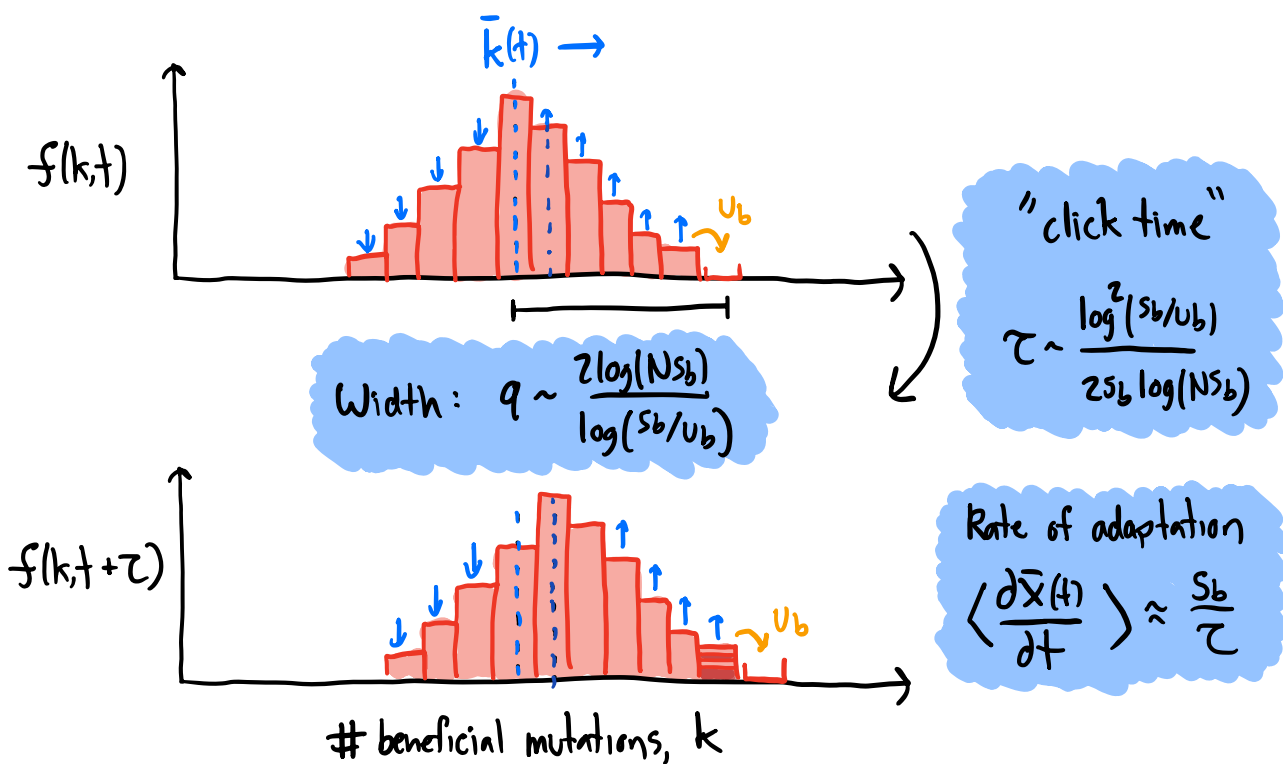
$$\tau_k \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b q^k}{U_b}\right)$$

Total contribution:

$$f_{\text{nose}}(t) \equiv \sum_{k=1}^{\infty} f_k(t) \equiv \frac{1}{Nq^s} e^{qs(t-\tau)} \Rightarrow$$

$$\tau \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b}\right)$$

⇒ Complete picture of dynamics of fitness dist'n:



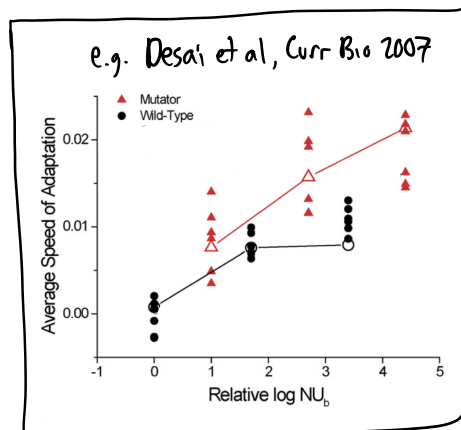
⇒ early tests for clonal interference in lab evolution experiments:

Successive mutations:

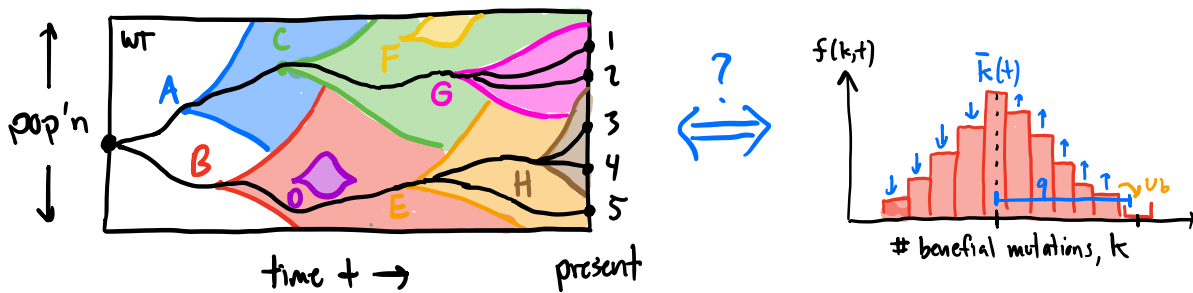
$$\langle \frac{d\bar{x}}{dt} \rangle \sim s_b^2 \cdot NU_b$$

clonal interference:

$$\langle \frac{d\bar{x}}{dt} \rangle \sim s_b^2 \cdot \frac{\log(Ns_b)}{\log^2(s_b/u_b)}$$

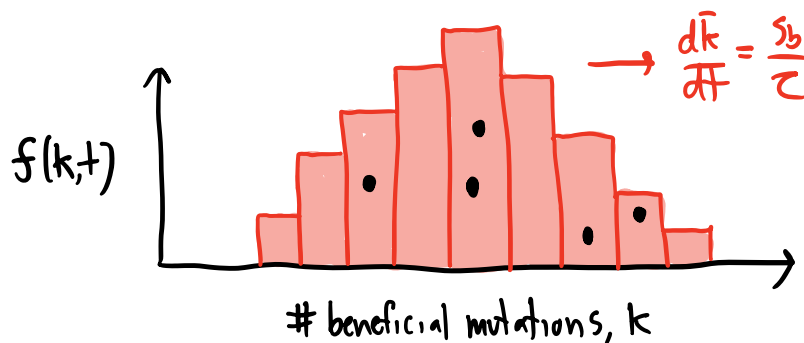


Next: Can we use this picture to understand genetic diversity backwards in time?

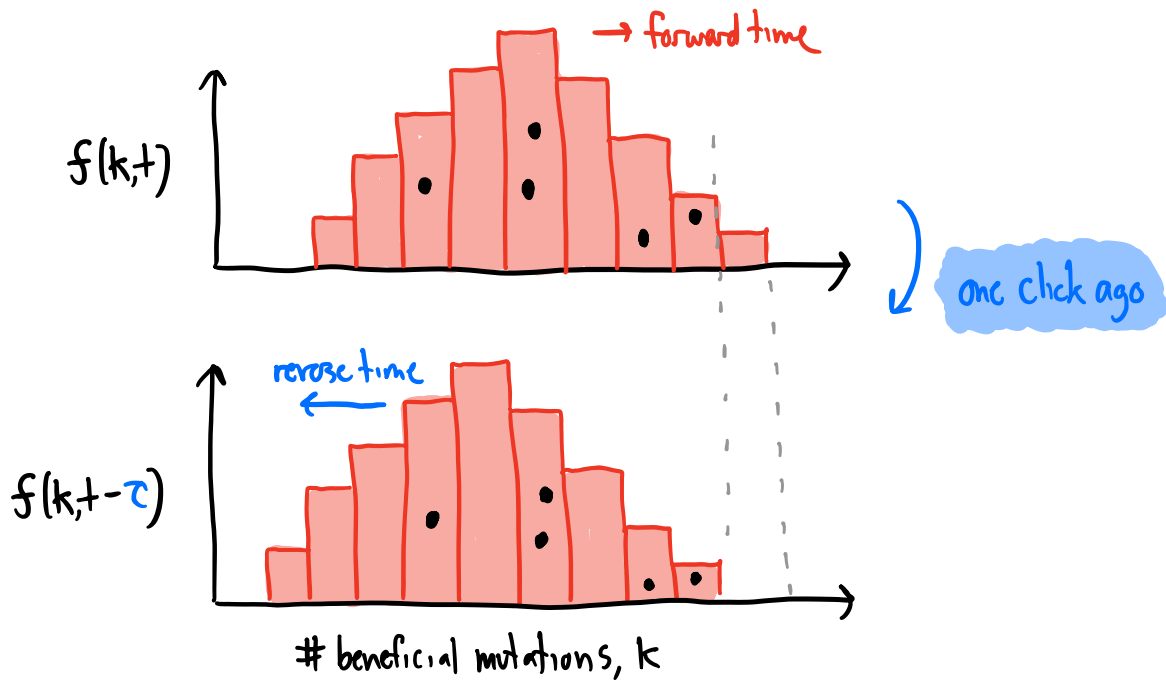


Answer: Yes we can! Let's start w/ some cartoons...

Step 1: draw sample of individuals from pop'n (present day)

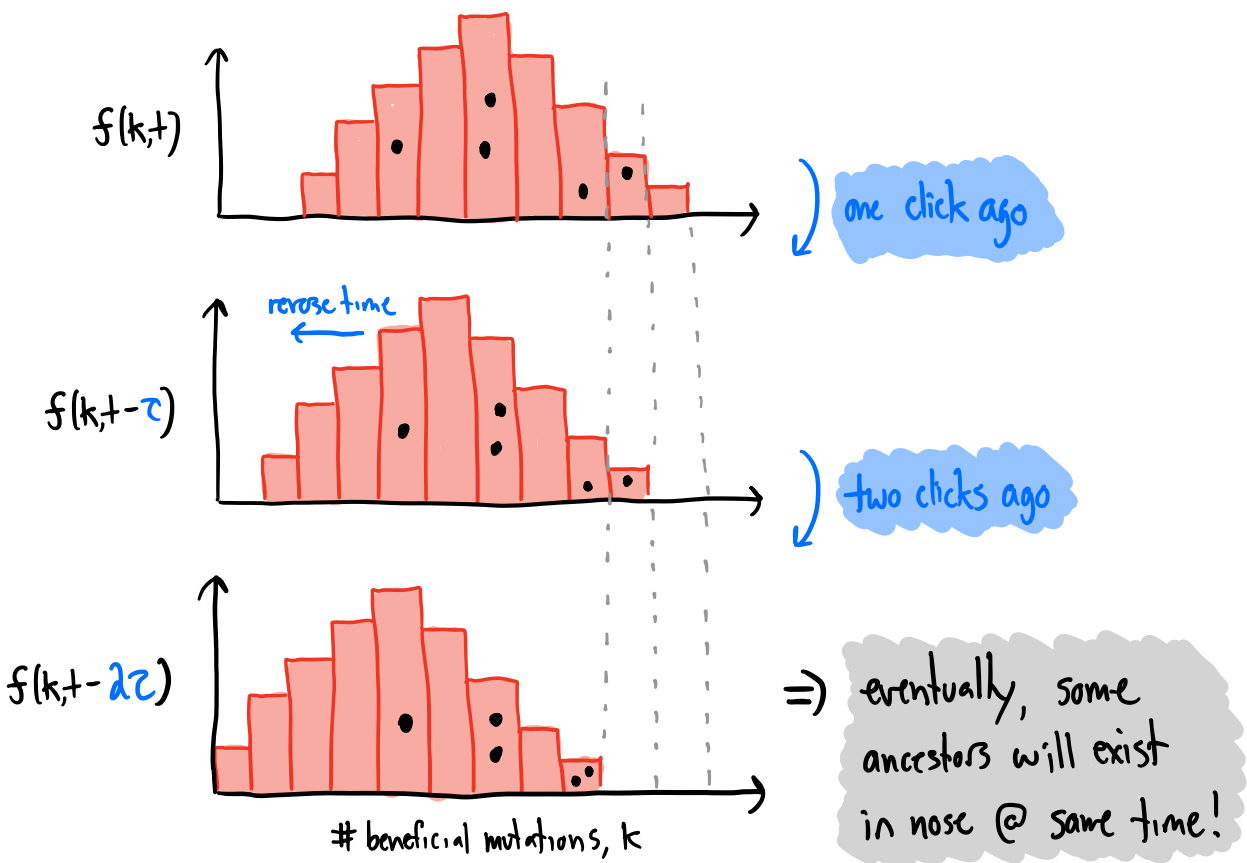


Step 2: where was everyone one click ago?



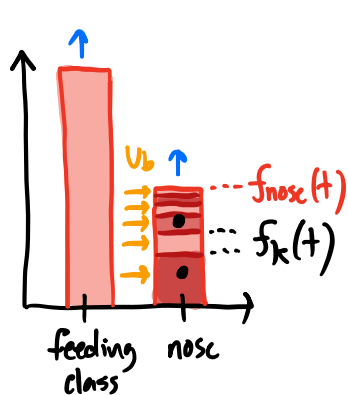
① can only coalesce if in same fitness class

② But little chance of coalescing in "bulk" of dist'n
(since $\tau \ll Nf_{q-1}(\tau), Nf_{q-2}(\tau), \text{etc.}$)



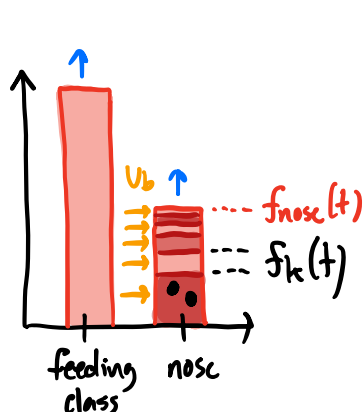
Two possible scenarios:

① Individuals are from separate lineages in the nose



↓↓
 separate ancestors
 in feeding class
 (distinct mut'n events)

② Individuals from same lineage in nose



Common ancestor
in feeding class

(coalescence w/in τ gens)

$$\Rightarrow \text{Probability: } p_c(2) = \sum_{k=1}^{\infty} \left(\frac{f_k(t)}{f_{nose}(t)} \right)^2 = \sum_{k=1}^{\infty} \left[\frac{\frac{1}{Nq s_b} e^{q s_b (t - \tau_k)}}{\frac{1}{Nq s_b} e^{q s_b (t - \tau)}} \right]^2$$

$$= \sum_{k=1}^{\infty} e^{-2 q s_b (\tau_k - \tau)}$$



only depends on establishment times τ_k !

⇒ if we plug-in typical values of τ_k & τ from heuristics:

$$\tau_k \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b q k}{u_b}\right); \quad \tau \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{u_b}\right)$$

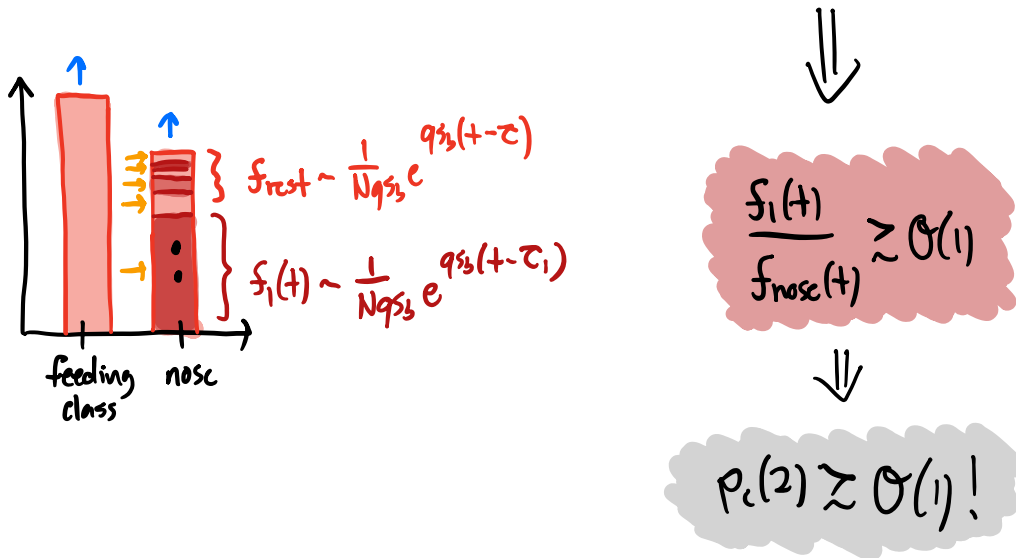
$$\Rightarrow p_c(z) = \sum_{k=1}^{\infty} e^{-2qs_b(\tau_k - \tau)} = \sum_{k=1}^{\infty} (qk)^{-\frac{2q}{q-1}} \approx \frac{1}{q^2}$$

~~⇒ suggests coalescence after $\sim q^2$ clicks ($T_{\text{thru}} \sim q^2 \tau$)~~

⇒ missing key part of puzzle: fluctuations

⇒ coalescence rare for typical lineage sizes,
but small chance of having anomalously early mutant
where coalescence is much more likely!

e.g. if first successful mutation occurs when $\tau_1 \approx \tau$:



\Rightarrow not a huge shift: typically, $\tau_1 - \tau \sim \frac{\log(q)}{q s_b} \ll \frac{1}{s_b} \ll \tau$

\Rightarrow occurs w/ total probability:

$$P_{\text{jackpot}} \sim \int_0^{\tau} d\tau_1 N U_b f_{q-1}(t) \cdot q s_b$$

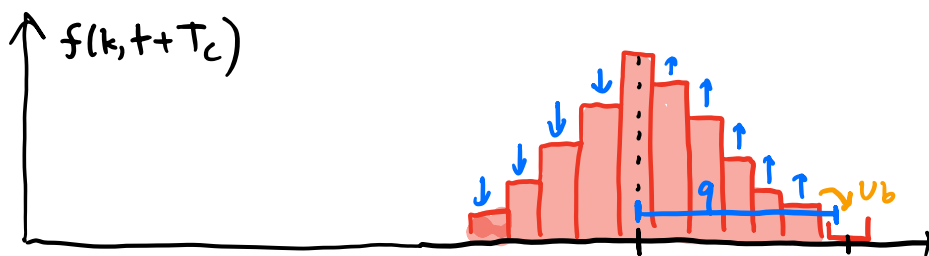
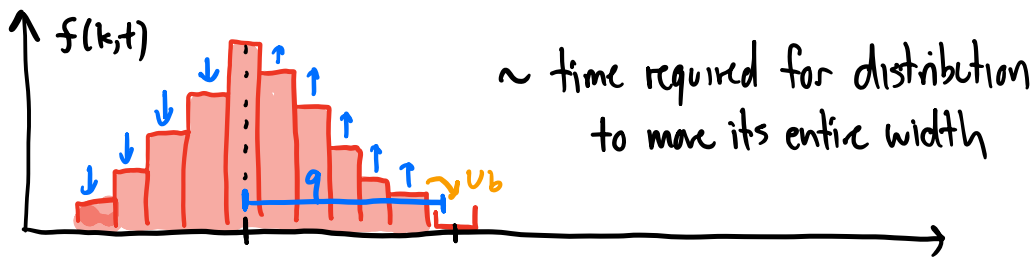
$$\sim \int_0^{\tau} d\tau_1 \cancel{N U_b} \cdot \frac{e^{(q-1)s_b t}}{\cancel{N q s_b}} \cdot \cancel{q s_b} \sim \frac{U_b}{(q-1)s_b} e^{(q-1)s_b \tau}$$

$$\sim \frac{1}{q-1} \approx \frac{1}{q}$$

$$\Rightarrow p_{\text{jackpot}} \sim \frac{1}{q} \quad (\gg \frac{1}{q^2})$$

\Rightarrow typical coalescence after $\frac{1}{p_{\text{jackpot}}} \sim q$ clicks

$$\Rightarrow \text{coalescent timescale } T_c \equiv q\tau \sim \frac{1}{s_b} \log\left(\frac{s_b}{u_b}\right)$$

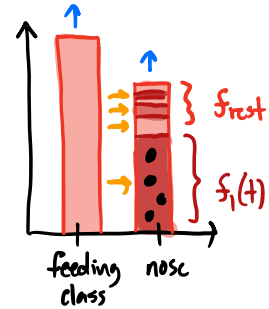


beneficial mutations, k

⇒ **fluctuations** were crucial for determining T_c !

⇒ coalescence is "bursty":

e.g. in larger sample size n :



$$P_c(n \rightarrow 1) \sim \left(\frac{f_1(t)}{f_1(t) + f_{rest}} \right)^n$$

⇒ $\mathcal{O}(1)$ if $f_{rest}(t)/f_1(t) \lesssim 1/n$

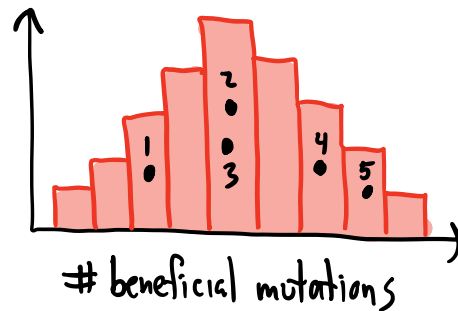
$$\Rightarrow P_{\text{jackpot}}(n) = \int_0^{\tau - \log(n)/q_5} d\tau_1 \, N U_b f_{q-1}(t) \cdot q_5 \sim 1/qn$$

⇒ i.e. multiple mergers likely! 

* For "formal" treatment, see Appendix C...

Another interesting feature of genealogies + travelling wave:

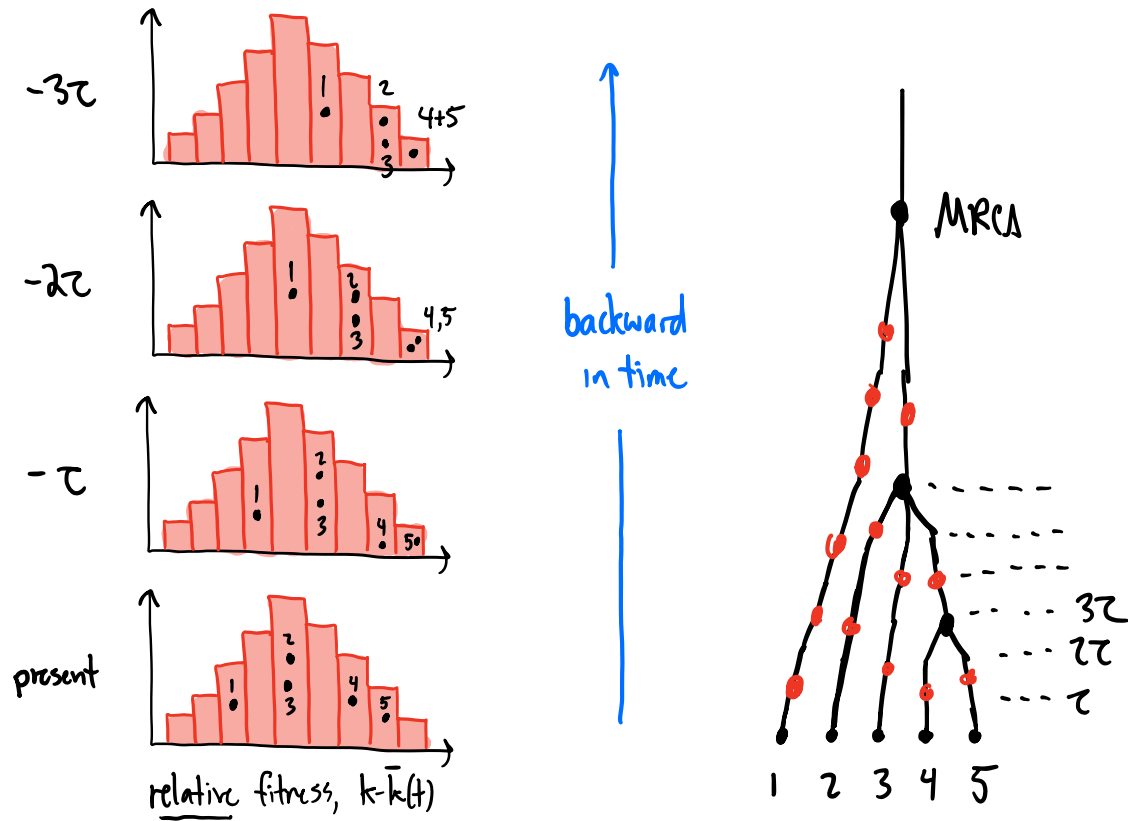
⇒ consider same example:



⇒ which individual's descendents are more likely to take over pop'n in future?

⇒ e.g. $5 \rightarrow 4 \rightarrow 2, 3 \rightarrow 1$

⇒ now let's try to "simulate" genealogy...



⇒ time (+burstiness) of coalescence in past

⇒ info about fitness in present

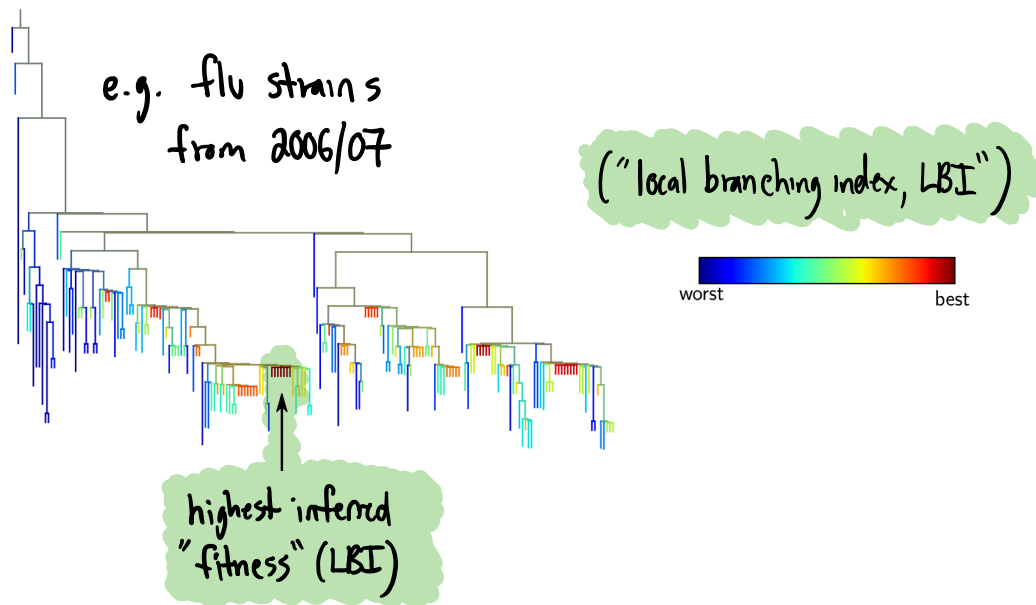
⇒ forecasts about who takes over in future!

Predicting evolution from the shape of genealogical trees

Richard A Neher^{1*}, Colin A Russell², Boris I Shraiman^{3*}

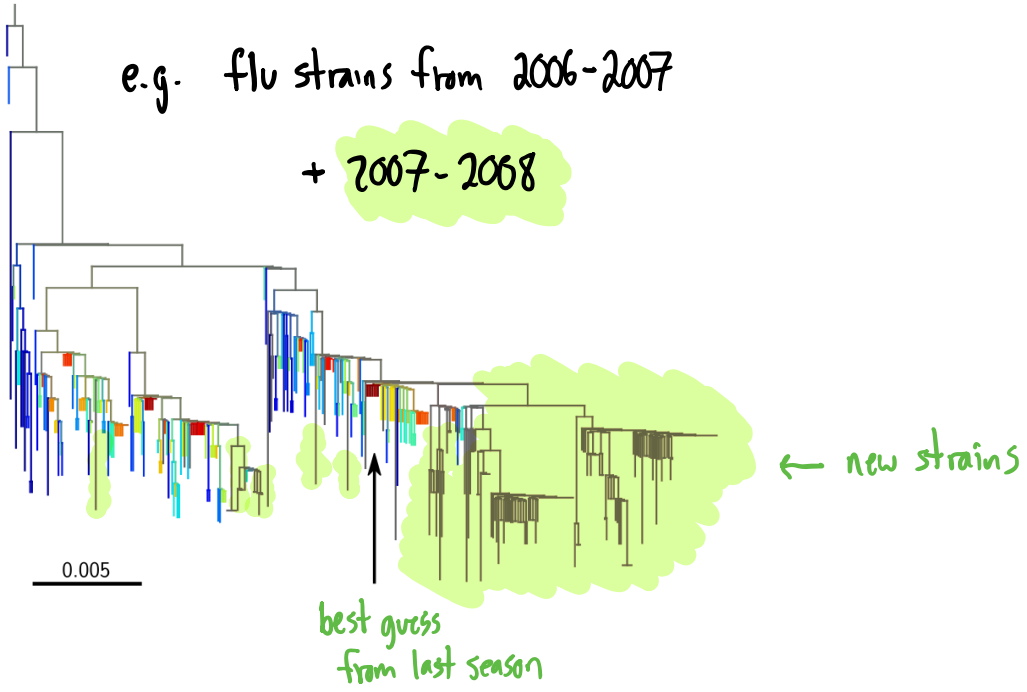
¹Evolutionary Dynamics and Biophysics, Max Planck Institute for Developmental Biology, Tübingen, Germany; ²Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom; ³Kavli Institute for Theoretical Physics, University of California, Santa Barbara, Santa Barbara, United States

⇒ implemented this idea for HA gene in influenza
(data from Problem #1 in HW1)

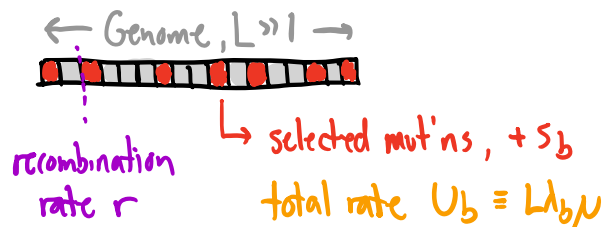


e.g. flu strains from 2006-2007

+ 2007-2008



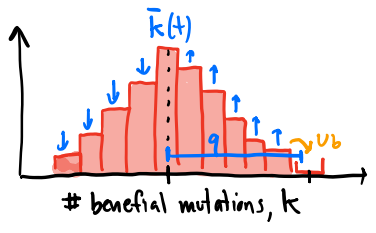
What about recombination?



$r = 0$



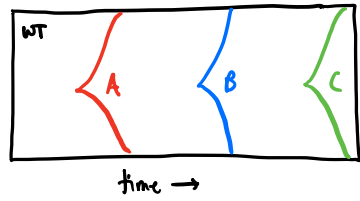
⇒



$r \gg N \cdot \mu_b \cdot s$



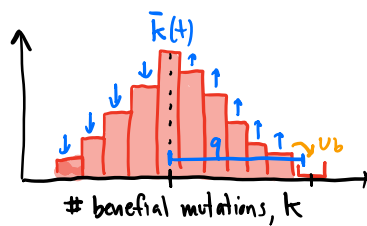
⇒



$r=0$



\Rightarrow



$U_{b,eff} = l^* \lambda_b \mu?$

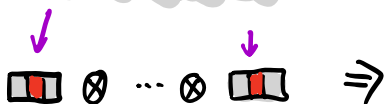
In between?



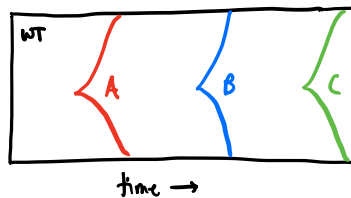
Blocks are ~ independent

... but multiple mut'n's / block!

$r \gg N \cdot \mu_b \cdot s$



\Rightarrow



If true, need:

① w/in blocks, recombination should be rare! ($r \approx 0$)

$\Rightarrow r l^* \cdot T_c(N, s_b, U_{eff}(l^*)) \ll 1$

② between blocks, recombination should be frequent!

$\Rightarrow r l^* \cdot T_c \gg 1$

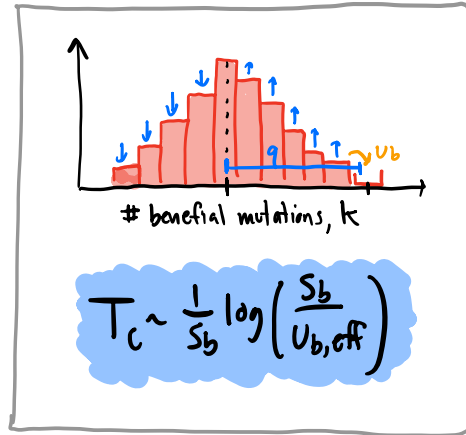
\Rightarrow can we (almost) satisfy both w/ $r l^* \cdot T_c \sim \mathcal{O}(1)$?

Linkage block ansatz



$$l^* \sim 1/rT_c$$

$$U_{b,eff} = l^* \lambda_b \mu$$



$$T_c \sim \frac{1}{S_b} \log\left(\frac{S_b}{U_{b,eff}}\right)$$

$$\Rightarrow \text{Self consistency: } T_c \sim \frac{1}{S_b} \log\left(\frac{S_b}{\mu \lambda_b} \cdot r T_c\right)$$

$$\Rightarrow \text{solution: } T_c \sim \frac{1}{S_b} \log\left(\frac{r}{\mu \lambda_b}\right)$$

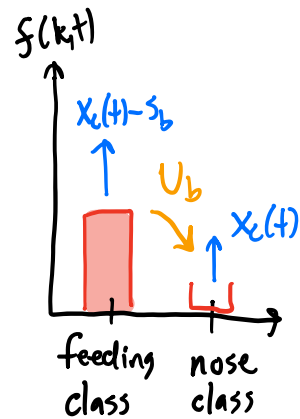
$$\Rightarrow l^* \sim \frac{S_b}{r} \log^{-1}\left(\frac{r}{\mu \lambda_b}\right)$$

$$\Rightarrow \text{self consistent if } N U_{b,eff} \log(N S_b) \gg 1 \quad \& \quad U_{eff,b} \ll S_b$$

$$\Rightarrow N \cdot \mu \lambda_b \cdot S_b \gg r \gg \mu \lambda_b$$

Appendix A: Formal analysis of the nose class

⇒ we can understand the establishment of the nose class more formally using the branching process framework that we studied in the 1st half of the course



⇒ Under our assumptions, nose can be described by LBP model:

$$\frac{df_q}{dt} = X_c(t) f_q + U_b f_{q-1}(t) + \sqrt{\frac{f_q}{N}} \eta(t)$$

w/ $f_q(0) = 0$ & time-varying:

selection: $X_c(t) = [q - \bar{k}(t)] s_b$

+
mutation: $U_b f_{q-1}(t) = \frac{U_b}{2Nq s_b} e^{\int_0^t (X_c(t) - s_b) dt}$

⇒ In their analysis, Desai & Fisher (2007) assumed that $\bar{k}(t) \approx 0$ throughout the establishment period, so that $X_c(t) \equiv qs_b$ & $f_{q-1}(t) = \frac{1}{Nqs} e^{(q-1)s_b t}$

⇒ Let's see how far we can get by relaxing this approx & explicitly modeling the "click" of $\bar{k}(t)$...

⇒ will be harder because time-varying fitness

$$X_c(t) = qs_b - s_b \bar{k}(t)$$

⇒ From our discussion in class, can take

$$\bar{k}(t) \equiv \frac{e^{s_b(t-t_c)}}{1 + e^{s_b(t-t_c)}}$$

where t_c is the time that $\bar{k}(t)$ clicks.

(later we will imagine that t_c is close to τ ...)

⇒ From SDE, the generating function $H_f(z, t) = \langle e^{-z \cdot f_q(t)} \rangle$
 satisfies the PDE:

$$\frac{\partial H_f}{\partial t} = \left[\chi_c(t) z - \frac{z^2}{2N} \right] \frac{\partial H_f}{\partial z} - z U_{bf_{q-1}}(t) H_f$$

w/ initial condition $H_f(z, 0) = 1$

⇒ can solve w/ method of characteristics:

$$\text{define: } \psi(t_R) = \log \left[H_f(z(t_R), t - t_R) \right]$$

$$\text{w/ } \psi(t) = 0, \quad z(0) = z,$$

$$\psi(0) \equiv \log H_f(z, t)$$

$$\Rightarrow \psi \text{ satisfies: } \frac{d\psi}{dt_R} = - \frac{\frac{\partial H_f}{\partial t}}{H_f} + \frac{\frac{\partial H_f}{\partial z}}{H_f} \left(\frac{dz}{dt_R} \right)$$

$$\Rightarrow \frac{d\psi}{dt_R} = \left\{ \frac{dz}{dt_R} - \left[\chi_c(t-t_R)z - \frac{z^2}{2N} \right] \right\} \frac{d \log H_f}{dz} + z(t_R) U_b f_{q-1}(t-t_R)$$

$$\Rightarrow \text{if } \frac{dz}{dt_R} = \chi_c(t-t_R)z - \frac{z^2}{2N} \quad \& \quad z(0) = z$$

$$\Rightarrow \psi(t_R) = \psi(0) + \int_0^{t_R} z(t_R) U_b f_{q-1}(t-t_R) dt_R'$$

$$\Rightarrow \log H_f(z, t) = - \int_0^t z(\tau) U_b f_{q-1}(t-\tau) d\tau$$

where

$$\frac{dz}{dt} = \chi_c(t-\tau)z - \frac{z^2}{2N}, \quad z(0) = z$$

\Rightarrow solution for characteristic curve is given by:

$$z(\tau) = \frac{z e^{\int_0^\tau \chi_c(t-\tau') d\tau'}}{1 + \frac{z}{2N} \int_0^\tau e^{\int_0^\tau \chi_c(t-\tau'') d\tau''} d\tau'}$$

(can plug in & check...)

$$\text{so } H_f(z, t) = \exp \left[- \int_0^t \frac{z \cdot U_b f_{q-1}(t-\tau) e^{\int_0^\tau x_c(t-\tau') d\tau'}}{1 + \frac{z}{zN} \int_0^\tau e^{\int_0^{\tau'} x_c(t-\tau'') d\tau''} d\tau'} d\tau \right]$$

$$= \exp \left[- \int_0^t \frac{z \cdot U_b f_{q-1}(u) e^{\int_u^+ x_c(u') du'}}{1 + \frac{z}{zN} \int_u^+ e^{\int_u^+ x_c(u'') du''} du'} du \right]$$

\Rightarrow again, helpful to define $v(t)$ s.t. $f_q(t) \equiv \frac{v(t)}{zNqsb} e^{\int_0^+ x_c(t') dt'}$

$$\Rightarrow H_v(z, t) \equiv \langle e^{-z \cdot v(t)} \rangle \equiv H_f \left(zNqsb e^{-\int_0^+ x_c(t') dt'} z, t \right)$$

$$\Rightarrow H_v(z, t) = \exp \left[- \int_0^+ \frac{z U_b f_{q-1}(u) zNqsb e^{-\int_0^u x_c(u') du'}}{1 + qsz \cdot \int_u^+ du' e^{-\int_0^{u'} x_c(u'') du''}} du \right]$$

\Rightarrow similar to single-locus case, we expect $v(t)$ to approach constant value v @ long times

$$\Rightarrow H_v(z) \equiv \lim_{t \rightarrow \infty} H_v(z, t)$$

$$\Rightarrow \log H_v(z) = - \int_0^{\infty} \frac{z \cdot U_b e^{-s_b t} dt}{1 + z \cdot q s_b \int_t^{\infty} dt' e^{-\int_0^{t'} x_c(t'') dt''}}$$

\Rightarrow Now we have to plug in our expression for $x_c(t)$:

$$x_c(t) = q s_b - \frac{s_b e^{s_b(t-t_c)}}{1 + e^{s_b(t-t_c)}}$$

$$\Rightarrow \int_0^t x_c(t') dt' = q s_b t - \log \left[\frac{1 + e^{s_b(t-t_c)}}{1 + e^{-s_b t_c}} \right]$$

$$\Rightarrow e^{-\int_0^t x_c(t') dt'} = e^{-q s_b t} \left[\frac{1 + e^{s_b(t-t_c)}}{1 + e^{-s_b t_c}} \right]$$

$$\Rightarrow q s_b \int_t^{\infty} dt' e^{-\int_0^{t'} x_c(t'') dt''} = \frac{e^{-q s_b t}}{1 + e^{-s_b t_c}} + \left(\frac{q}{q-1} \right) \frac{e^{-q s_b t} e^{s_b(t-t_c)}}{1 + e^{-s_b t_c}}$$

and hence:

$$\log H_V(z) = - \int_0^{\infty} \frac{z \cdot U_b e^{-s_b t} dt}{1 + z \cdot e^{-q s_b t} \left[1 + e^{s_b(t-t_c)} \left(\frac{q}{q-1} \right) \right]}$$

(where we have assumed that the click time t_c is $\gg \frac{1}{s_b}$)

\Rightarrow for large q + relevant values of z , this integral will be dominated by times w/in $o(\frac{1}{s_b})$ of τ .

\Rightarrow can extend lower limit of integral to $t = -\infty$
w/o much error...

\Rightarrow if t_c is also w/in $o(\frac{1}{s_b})$ of τ , we can expand $e^{s_b(t-t_c)}$ term in denominator, so that

$$\log H_V(z) \approx - \int_{-\infty}^{\infty} \frac{z \cdot U_b e^{-s_b t} dt}{1 + z \cdot z \cdot e^{-q s_b t}}$$

changing variables to $\xi = (2z)^{1/q} e^{-s_b t}$,

$$\log H_v(z) = \exp \left[-\frac{U_b}{S_b} \cdot z^{1-1/q} \cdot \left(2^{-1/q} \int_0^\infty \frac{\xi d\xi}{1+\xi^q} \right) \right] \xrightarrow{+O(1/q)}$$

$$\Rightarrow H_v(z) = e^{-\frac{U_b}{S_b} z^{1-1/q}}$$

\Rightarrow typical value of v occurs when $H_v(z = \frac{1}{v^*}) = e^{-1}$

$$\Rightarrow v^* = \left(\frac{S_b}{U_b} \right)^{\frac{q}{q-1}}$$

$$\Rightarrow \text{substituting into } f_q(t) = \frac{v}{2NqS} e^{qst} \equiv \frac{e^{qs(t-\tau)}}{2NqS}$$

$$\Rightarrow \text{typical value of } f_q^*(t) = \frac{e^{s_b t}}{NqS} \left(\frac{S_b}{U_b} \right)^{\frac{q}{q-1}}$$

⇒ typical value of establishment time:

$$\tau^* = \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{u_b}\right)$$

⇒ consistent w/ results from
simpler heuristic argument!

Appendix B: How many lineages contribute to new nose?

Recall in heuristic argument, we had:

$$f_{\text{nose}}(t) \equiv \sum_{k=1}^{K_{\text{max}}} f_k(t) = \frac{1}{Nq_s} e^{qs_b(t-\tau)} \cdot \sum_{k=1}^{K_{\text{max}}} \frac{1}{q} k^{1+1/q}$$

+ argued that sum over k converged to ≈ 1 .

\Rightarrow Let's look @ this more carefully...

\Rightarrow if $K_{\text{max}} \gg 1$ (will revisit below)

$$\Rightarrow \sum_{k=1}^{K_{\text{max}}} \frac{1}{q} \frac{1}{k^{1+1/q}} \approx \int_1^{K_{\text{max}}} \frac{dk}{q k^{1+1/q}} = 1 - e^{-\frac{1}{q} \log K_{\text{max}}}$$

Thus, sum converges to 1 provided that

$\log K_{\text{max}}$ is large compared to q

\Rightarrow how does this translate to establishment times τ_k ?

recall that $\tau_k - \tau_1 \sim \frac{1}{q s_b} \log(k)$, so condition becomes:

$$\Rightarrow \tau_{K_{\max}} - \tau_1 \sim \frac{1}{q s_b} \log(K_{\max}) \gg \frac{1}{s_b}$$

Thus, mutations that establish $\gg \frac{1}{s_b}$ after τ_1
have negligible contribution to $f_{\text{nose}}(t)$, τ , etc

\Rightarrow since $\frac{1}{s_b} \ll \tau$, this happens long before next click.

\Rightarrow can take $K_{\max} \approx \infty$ w/o losing any accuracy

i.e., can pretend that infinite # of muts
contribute to establishment of new nose.

Appendix C: formal analysis of coalescence in the nose

Recall: main result for stochastic size of nose:

$$f_{\text{nose}}(t) \equiv \frac{\nu}{2Nq s_b} e^{q s_b t} \Rightarrow H_{\nu}(z) \approx e^{-\frac{\nu}{s_b} z^{1-1/q}}$$

(supplement of lecture 19)

Let's fine-grain this further:

$$\Rightarrow \text{let } f_{\ell}(t) \equiv \text{freq of lineage in nose founded by beneficial mutation @ site } \ell$$

$$\Rightarrow \text{then } H_{\nu_{\ell}}(z) \approx e^{-\frac{\nu}{s_b} z^{1-1/q}}$$

\Rightarrow Probability that 2 individuals coalesce =
probability that they came from same lineage:

$$\Rightarrow \rho_c(z) = \left\langle \sum_{l=1}^{L_b} \left(\frac{f_l(t)}{\sum_{e'} f_{e'}(t)} \right)^z \right\rangle = \left\langle \sum_{e=1}^{L_b} \left(\frac{\nu_e}{\sum_{e'} \nu_{e'}} \right)^z \right\rangle$$

Trick: using $\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z} dz = 1$, can write as

$$\rho_c(z) = \left\langle \sum_e \left(\frac{\nu_e}{\sum_{e'} \nu_{e'}} \right)^z \right\rangle = \left\langle \sum_e \nu_e^z \int_0^\infty dz z e^{-\left(\sum_{e'} \nu_{e'}\right) z} \right\rangle$$

$$= \sum_e \left\langle \int_0^\infty dz z \cdot \left(\nu_e^z e^{-z \nu_e} \right) \cdot \prod_{e' \neq e} e^{-\nu_{e'} z} \right\rangle$$

$$= \sum_e \int_0^\infty dz z \cdot \frac{\partial H_{\nu_e}(z)}{\partial z^2} \cdot \prod_{e' \neq e} H_{\nu_{e'}}(z)$$

\Rightarrow using results above for $H_{\nu_e}(z)$ and $H_\nu(z)$,

$$\Rightarrow \frac{\partial H_{\nu_e}}{\partial z} = -\frac{\mu_e}{s_b} (1 - \frac{1}{q}) z^{-\frac{1}{q}} H_{\nu_e}(z)$$

$$\Rightarrow \frac{\partial^2 H_{\nu_\ell}}{\partial z^2} = \frac{1}{q} \frac{\mu_\ell}{s_b} \left(1 - \frac{1}{q}\right) z^{-1 - \frac{1}{q}} H_{\nu_\ell}(z) + \mathcal{O}(\mu_\ell^2)$$

so that:

$$\begin{aligned} p_c(z) &= \sum_\ell \int_0^\infty dz \cdot z \cdot \frac{\partial^2 H_{\nu_\ell}(z)}{\partial z^2} \cdot \prod_{\ell' \neq \ell} H_{\nu_{\ell'}}(z) \\ &= \frac{1}{q} \sum_\ell \frac{\mu_\ell}{\nu_b} \int_0^\infty dz \frac{\nu_b}{s_b} \left(1 - \frac{1}{q}\right) z^{-\frac{1}{q}} H_{\nu_\ell}(z) \stackrel{=}{=} \prod_\ell H_{\nu_\ell}(z) \\ &= \frac{1}{q} \int_0^\infty - \frac{\partial H_\nu(z)}{\partial z} = \frac{1}{q} \left[H_\nu \Big|_0^1 - H_\nu \Big|_\infty^0 \right] \\ &= \frac{1}{q} \quad \text{as desired} \end{aligned}$$

Can do same thing for larger samples:

$$p_c(n) = \left\langle \sum_\ell \left(\frac{\nu_\ell}{\sum_{e'} \nu_{e'}} \right)^n \right\rangle = \sum_\ell \int_0^\infty dz \cdot \frac{(-1)^n z}{\Gamma(n)} \frac{\partial^n H_{\nu_\ell}(z)}{\partial z^n} \prod_{\ell' \neq \ell} H_{\nu_{\ell'}}(z)$$

$$\Rightarrow \frac{d^n H_{\nu_e}(z)}{dz^n} = \frac{N_e}{S_b} \frac{(-1)^n}{q} \left(1 - \frac{1}{q}\right) \prod_{k=1}^{n-2} \left(k + \frac{1}{q}\right) z^{-n+1-\frac{1}{q}} H_{\nu_e}(z) + \mathcal{O}(N_e z)$$

$$\approx \frac{N_e}{S_b} \frac{(-1)^n (n-2)!}{q} z^{-n+1-\frac{1}{q}} H_{\nu_e}(z)$$

$$\Rightarrow \rho_c(n) = \frac{1}{q} \frac{\Gamma(n-1)}{\Gamma(n)} \cdot \sum_e \frac{N_e}{U_b} \int_0^\infty dz \cdot \frac{-dH_{\nu_e}(z)}{dz}$$

$$\approx \frac{1}{q(n-1)}$$