Discrete transitions, Markov process

discrete states labeled by $n \in \{0, 1, 2, \dots\}$

in an infinitessimal time Δt there is a probability states change

probabilities of transitioning = rates × Δt:

can depend on initial state

$$p_{m,n} = r^{(m)}_{m,n}\,\Delta t, \qquad n \neq m \qquad\qquad p_{m,m} = \Big(1 - \big(\sum_{n\neq m} r_{m,n}\big)\,\Delta t\Big)$$

initial final

Probability system is in state n at time $t+\Delta t$:

$$P(n, t+\Delta t) = P(n,t)\,\underbrace{p_{n,n}}_{1 - \sum_{n'\neq n} r_{n,n'}\,\Delta t} + \sum_{n'\neq n} P(n',t)\,p_{n',n}$$

$$\frac{P(n,t+\Delta t) - P(n,t)}{\Delta t} = -P(n,t)\sum_{n'\neq n} r_{n,n'}\,\cancel{\Delta t} + \sum_{n'\neq n} P(n',t)\,r_{n',n}\,\cancel{\Delta t}$$

$$\frac{\partial P(n,t)}{\partial t} = \underbrace{\sum_{n'\neq n} r_{n',n}\,P(n',t)}_{\text{probability flux into } n} - \underbrace{P(n,t)\sum_{n'\neq n} r_{n,n'}}_{\text{probability flux out of } n}$$

Birth death process

n = number of individuals

$p_{n,n+1}$: prob. of transitioning from n to $n+1$ in Δt $= n\,r(n)\,\Delta t$

n independent individuals

birth rate may be a function of n — resources

$$p_{n,n-1} = n\,\mu(n)\,\Delta t$$

$$p_{n,n} = 1 - (r+\mu)\,n\,\Delta t$$

$$P(n, t+\Delta t) = P(n,t)\underbrace{p_{nn}}_{1-n(r+\mu)\Delta t} + P(n+1,t)\overbrace{p_{n+1,n}}^{(n+1)\mu(n+1)\Delta t} + P(n-1,t)\underbrace{p_{n-1,n}}_{(n-1)r(n-1)\Delta t}$$

Forward Master eqn for birth-death process

$$\Rightarrow \boxed{\frac{\partial P(n,t)}{\partial t} = (n-1)\,r\,P(n-1,t) - (r+\mu)\,n\,P(n,t) + (n+1)\,\mu\,P(n+1,t)}$$

$\Rightarrow$ 2-term recursion

need B.C. for $\underline{n=0}$:

Boundary condition at $n=0$:

$$\boxed{\frac{\partial P(0,t)}{\partial t} = \mu P(1,t)},$$

$n=0$ is an absorbing boundary condition (state)

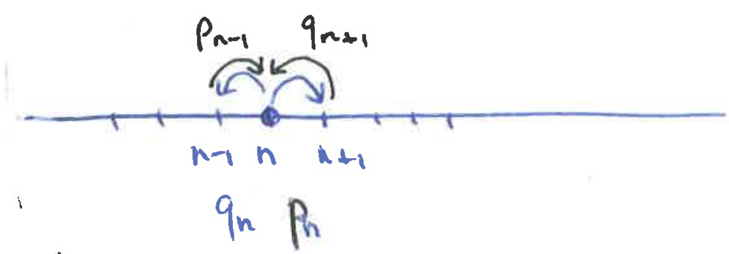
because $n\gamma$ and $n\mu$,

Nothing can happen if $n=0$

Initial condition $\boxed{P(n,0) = \mathbb{1}(n,N)}$ starting with N particles

---

Analogous to a 1-D discrete random walk:



$$P(n,t+\Delta t) = P(n,t)\left[1-(p_n+q_n)\Delta t\right]$$
$$+\ q_{n+1}P(n+1,t)$$
$$+\ p_{n-1}P(n-1,t)$$

p, q are forward, backward hopping rates

$$\boxed{\frac{\partial P(n,t)}{\partial t} = p_{n-1}P(n-1,t) - (p_n+q_n)P(n,t) + q_{n+1}P(n+1,t)}$$

(can be unbounded)

Solve birth-death Master eqn using generating functions (analogous to Laplace transform). This only works if μ and r are generally constants:

$$G(z,t) \equiv \sum_{n=0}^{\infty} P(n,t)\, z^n$$

Multiply Master Eq by z^n and Sum:

$$\sum_{n=0}^{\infty} \frac{\partial P(n,t)}{\partial t} z^n \equiv \frac{\partial G(z,t)}{\partial t}$$

$$-\sum_{n=0}^{\infty} (r+\mu)\, n\, P(n,t)\, z^n$$

$$= -(r+\mu)\, z \frac{\partial}{\partial z} \sum_{n=0}^{\infty} P(n,t)\, z^n = -(r+\mu)\, z \frac{\partial G}{\partial z}$$

$$z \sum_{n=1}^{\infty} r\,(n-1)\, P(n-1,t)\, z^{n-1} = r z^2 \frac{\partial G}{\partial z}$$

$$\frac{\mu}{z} \sum_{n=0}^{\infty} P(n+1,t)\, z^{n+1}\, (n+1) = \frac{\mu}{z} z \frac{\partial}{\partial z}\left[G(z) - P_0\right] = \mu \frac{\partial G}{\partial z}$$

$$\boxed{\begin{aligned} \frac{\partial G(z,t)}{\partial t} &= \left[r z^2 - (\mu+r) z + \mu\right] \frac{\partial G}{\partial z} \\ \sum_{n=0}^{\infty} \mathbb{1}(n,N)\, z^n &= z^N \equiv G(z,0) \end{aligned}}$$

Solve using Method of characteristics: along $\frac{dz(t)}{dt} = -\left(r z^2 - (\mu+r) z + \mu\right)$,

$\frac{dG}{dt} = 0$

$$\Rightarrow \quad G(z,t) = \left[\frac{e^{-(r-\mu)t}(rz-\mu) - \mu(z-1)}{e^{-(r-\mu)t}(rz-\mu) - r(z-1)}\right]^N$$

$$\xrightarrow[\mu=r]{} \left[\frac{1-(rt-1)(z-1)}{1-rt(z-1)}\right]^N$$

Check $t=0$

$$\frac{z_+(z-z_-) - z_- z + z_- z_+}{z - z_- - z + z_+} = \frac{z_+ z - z_- z}{z_+ - z_-} = z \quad ✓$$

$$G(1,t) = \left[\frac{z_+(1-z_-)e^{-\gamma t} - z_-(1-z_+)}{(1-z_-)e^{-\gamma t} - (1-z_+)}\right]^N$$

$$z_\pm = \frac{1+\bar{\mu}}{2} \pm \frac{1}{2}\sqrt{(1+\bar{\mu})^2 - 4\bar{\mu}}$$

$$\frac{1+\bar{\mu}}{2} \pm \frac{1}{2}\sqrt{(1-\bar{\mu})^2} = \frac{1+\bar{\mu}}{2} \pm \frac{1}{2}(1-\bar{\mu}) = 1,\ \bar{\mu}$$

$$G(1,t) = 1 \quad ✓$$

$$P(0,t) = \left[\frac{-z_+ z_- e^{-\gamma t} + z_+ z_-}{-z_- e^{-\gamma t} + z_+}\right]^N = \left[\frac{\mu(1-e^{-(r-\mu)t})}{r - \mu e^{-(r-\mu)t}}\right]^N \quad ✓$$

$$G(z,t) = P(0,t) + z P(1,t) + z^2 P(2,t) + \ldots$$

take power series about $z=0$ to find $P(n,t)$

$$P(n,t) = \frac{1}{n!} \frac{\partial^n G}{\partial z^n}\Big|_{z=0} ,$$ or, take inverse z transform:

$$P(n,t) = \oint \frac{G(z)}{z^{n+1}} \frac{dz}{2\pi i}$$

$$\boxed{\begin{aligned} P(0,t) &= \left[\frac{\mu - \mu e^{-(r-\mu)t}}{r - \mu e^{-(r-\mu)t}}\right]^N \qquad r \neq \mu \\ &= \left(\frac{rt}{1+rt}\right)^N \qquad r = \mu \end{aligned}}$$

Mean $\langle n \rangle = \sum_{n=0}^{\infty} n P(n,t) = \frac{\partial}{\partial z} \sum P(n,t) z^n \Big|_{z=1} = \frac{\partial G}{\partial z}\Big|_{z=1}$

$$\Rightarrow \langle \dot{n} \rangle = (r-\mu) \langle n \rangle$$

$$\boxed{\langle n \rangle = N e^{(r-\mu)t}}$$

S.D. $$\boxed{\langle n^2 \rangle - \langle n \rangle^2 = N \frac{(r+\mu)}{(r-\mu)} e^{(r-\mu)t} \left(e^{(r-\mu)t} - 1\right)}$$

Probability of extinction

$$\lim_{t\to\infty} P(0,t) = \begin{cases} 1 & r \le \mu \quad \text{certain extinction} \\ \left(\frac{\mu}{r}\right)^N & r > \mu \quad \text{probability of runaway population explosion} \end{cases}$$

Time to exctinction $(r \le \mu)$

probability that extinction occured in $[t, t+dt]$

$$[P(0,t+\Delta t) - P(0,t)] \equiv \overbrace{W(t)}^{\text{extinction time distribution}} \Delta t$$

where $W(t) = \frac{\partial P(0,t)}{\partial t} = -\frac{\partial S(t)}{\partial t}$ where

"Survival" probability $S(t) \equiv 1 - P(0,t)$

Mean time to extinction $\langle T \rangle \equiv \int_0^\infty W(t)\, t\, dt$

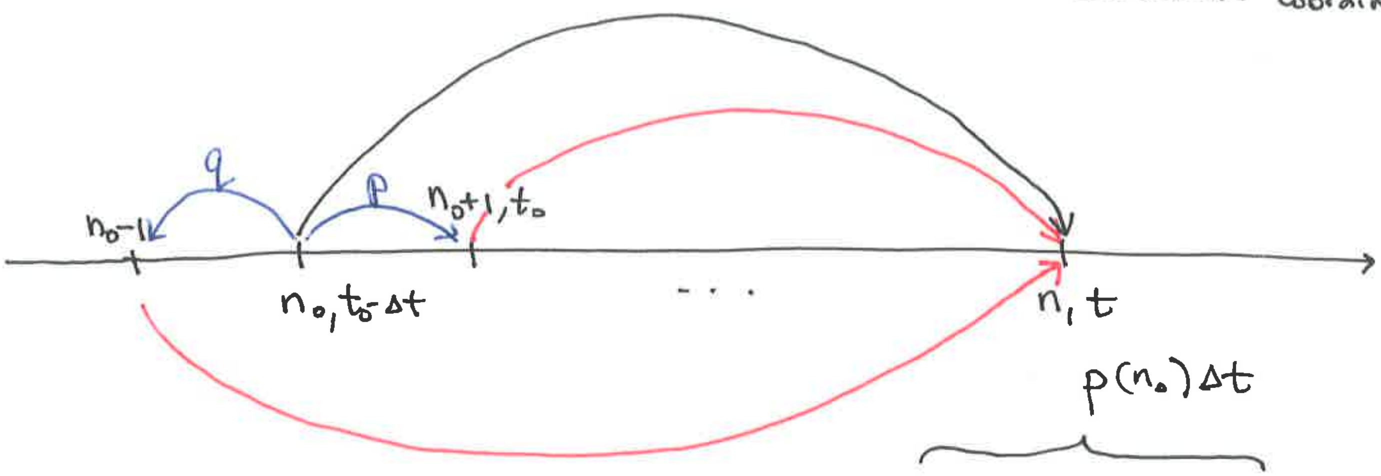
$$= -\int_0^\infty \left(\frac{\partial S}{\partial t}\right) t\, dt = t\,S(t)\Big|_0^\infty + \int_0^\infty S(t)\, dt$$

If $r > \mu$, conditional extinction time distribution

$$W(t\,|\,\text{ext.})\, dt = \frac{1}{P(0,\infty)} \left(-\frac{dS}{dt}\right)$$

Can find $\langle T \rangle$ or $S(t)$ directly using "Backward Eqn"

Consider a general 1-D random walk and $P(\underbrace{n,t}_{\text{final coordinate}} \mid \underbrace{n_0, t_0 - \Delta t}_{\text{Starting coordinate}})$



$$
\begin{aligned}
P(n,t \mid n_0, t_0 - \Delta t) = \; & P(n,t \mid n_0+1, t_0)\, \overbrace{P(n_0+1, t_0 \mid n_0, t_0 - \Delta t)}^{p(n_0)\Delta t} \\
& + P(n,t \mid n_0-1, t_0)\, \underbrace{P(n_0-1, t_0 \mid n_0, t_0 - \Delta t)}_{q(n_0)\Delta t} \\
& + P(n,t \mid n_0, t_0)\, \underbrace{P(n_0, t_0 \mid n_0, t_0 - \Delta t)}_{\substack{\text{prob of staying} \\ 1-(p(n_0)+q(n_0))\Delta t}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{P(n,t \mid n_0, t_0 - \Delta t) - P(n,t \mid n_0, t_0)}{\Delta t} = \; & q(n_0)\, P(n,t \mid n_0 - 1, t_0) \\
& - \left(p(n_0) + q(n_0)\right) P(n,t \mid n_0, t_0) \\
& + p(n_0)\, P(n,t \mid n_0 + 1, t_0) \\
= & -\frac{\partial P(n,t \mid n_0, t_0)}{\partial t_0}
\end{aligned}
$$

$$-\frac{\partial P(n,t|n_0,t_0)}{\partial t_0} = q\,P(n,t|n_0-1,t_0) - (p+q)\,P(n,t|n_0,t_0) + p\,P(n,t|n_0+1,t_0)$$

Recursion in n_0 and derivative in t_0, the starting time.

We can easily define a set of configurations $\mathcal{N}$

$$\sum_n^{\mathcal{N}} P(n,t|n_0,t_0) \equiv S(t|n_0,t_0) \equiv \text{survival within } \mathcal{N} \text{ up to time } t \text{ given system was at state } n_0 \text{ at time } t_0$$

<u>Initial condition</u> $S(t_0|n_0,t_0) = 1$

<u>Boundary condition</u> $S(t>t_0|n_0 \in \partial\mathcal{N}, t_0) = 0$

For time-inhomogeneous process,

$-\dfrac{\partial P}{\partial t_0} = +\dfrac{\partial P}{\partial t}$, so summing the Backward Eqn (can do this because it ~~is independent of n~~) depends on $P(n,t|$

$$\boxed{\frac{\partial S(t|n_0,t_0)}{\partial t} = q\,S(t|n_0-1,t_0) - (p+q)\,S(t|n_0,t_0) + p\,S(t|n_0+1,t_0)}$$

$\Rightarrow$ Solve directly for survival probability

integrate directly

$$\int_0^\infty \frac{\partial S}{\partial t}\,dt = \underbrace{S(\infty|n_0,t_0)}_{0} - \underbrace{S(0|n_0,t_0)}_{1} = q\,T(n_0-1) - (p+q)\,T(n_0) + p\,T(n_0+1)$$

$\Rightarrow$ recursion relation for $T(n_0)$ $\left(p\nabla^2 T \;\; +(p-q)\nabla T = -1\right)$

Consider other simple stochastic process (no spatial dependence — "well-mixed")

* regulated birth $\quad X \xrightarrow{r_0(1-n/K)} 2X, \quad X \xrightarrow{\mu_0} \phi$

* regulated death $\quad X \xrightarrow{r_0} 2X, \quad X \xrightarrow{\mu_0\left(1+\frac{r_0 n}{\mu_0 K}\right)} \phi$

* annihilation $\quad X \xrightarrow[r_0-\mu_0-\frac{r_0}{K}]{} 2X, \quad X+X \xrightarrow[\frac{r_0}{K}]{} \phi$

---

* regulated birth, birth rate depends on current population (consumption of resources)

K = carrying capacity

$$\dot{P}_n(t) = r_0\left(1-\frac{n-1}{K}\right)(n-1)P_{n-1} - \left(r_0\left(1-\frac{n}{K}\right)+\mu_0\right)n P_n$$

$$+ \mu_0 (n+1) P_{n+1}$$

$$\times \sum_{n=0}^{\infty} n \dot{P}_n(t) \Rightarrow \boxed{\langle \dot{n} \rangle = (r_0-\mu_0)\langle n \rangle - \frac{r_0}{K}\langle n^2 \rangle}$$

$$r_0 \sum_{n=0}^{\infty} n(n-1) P_{n-1} = r_0 \sum_{m=0}^{\infty} (m+1) m P_m$$

$$= r_0 \langle n(n+1) \rangle$$

$$\sum_{n=0}^{\infty} n^2 \dot{P}_n(t) = \boxed{\frac{d\langle n^2 \rangle}{dt} = 2r_0\langle n^2 \rangle + r_0\langle n \rangle - 2\mu_0\langle n^2 \rangle + \mu_0\langle n \rangle - \frac{2r_0}{K}\langle n^3 \rangle - \frac{r_0}{K}\langle n^2 \rangle}$$

* Regulated death

$$\dot{P}_n(t) = r_0(n-1)P_{n-1} - \left(r_0 + \mu_0\left(1+\frac{r_0 n}{\mu_0 K}\right)\right) n P_n + \mu_0\left(1+\frac{r_0}{\mu_0}\frac{n+1}{K}\right)(n+1)P_{n+1}$$

$$\boxed{\langle \dot{n} \rangle = (r_0-\mu_0)\langle n \rangle - \frac{r_0}{K}\langle n^2 \rangle}$$

$$\boxed{\frac{d\langle n^2 \rangle}{dt} = 2(r_0-\mu_0)\langle n^2 \rangle + (r_0+\mu_0)\langle n \rangle - \frac{2r_0}{K}\langle n^3 \rangle + \frac{r_0}{K}\langle n^2 \rangle}$$

* Annihilation

$$\underset{r_0-\mu_0-\frac{r_0}{K}}{X \to 2X} \qquad \overbrace{X+X}^{2 \text{ particles}} \xrightarrow[\frac{r_0}{K}]{} \phi \quad \text{0 particles}$$

$$\dot{P}_n(t) = \underbrace{\frac{r_0}{2K}}_{} \left[(n+1)(n+2)P_{n+2} - n(n-1)P_n\right]$$

½ from indistinguishability

$$+ \left(r_0-\mu_0-\frac{r_0}{K}\right)\left[(n-1)P_{n-1} - nP_n\right]$$

$$\sum_{n=0}^{\infty} n\dot{P}_n(t) = \frac{r_0}{2K}\sum_{n=0}^{\infty} n(n+1)(n+2)P_{n+2} + \dots$$

$$= \frac{r_0}{2K}\sum_{m=0}^{\infty} m(m-1)(m-2)P_m + \dots$$

$$= \frac{r_0}{2K}\langle n(n-1)(n-2) \rangle + \dots$$

$$\boxed{\langle \dot{n} \rangle = (r_0-\mu_0)\langle n \rangle - \frac{r_0}{K}\langle n^2 \rangle}$$

$$\frac{d\langle n^2\rangle}{dt} = 2(r_0-\mu_0)\langle n^2\rangle + (r_0-\mu_0)\langle n\rangle - \frac{3r_0}{K}\langle n\rangle + \frac{2r_0}{K}\langle n^3\rangle + \frac{2r_0}{K}\langle n^2\rangle$$

---

* Regulated birth: $\langle \dot{n}\rangle = (r_0-\mu_0)\langle n\rangle - \frac{r_0}{K}\langle n^2\rangle$

$$\langle \dot{n^2}\rangle = 2\left(r_0-\mu_0 - \frac{r_0}{2K}\right)\langle n^2\rangle + (r_0+\mu_0)\langle n\rangle - \frac{2r_0}{K}\langle n^3\rangle$$

* Regulated death: $\langle \dot{n}\rangle = (r_0-\mu_0)\langle n\rangle - \frac{r_0}{K}\langle n^2\rangle$

$$\langle \dot{n^2}\rangle = 2\left(r_0-\mu_0 + \frac{r_0}{2K}\right)\langle n^2\rangle + (r_0+\mu_0)\langle n\rangle - \frac{2r_0}{K}\langle n^3\rangle$$

* Annihilation: $\langle \dot{n}\rangle = (r_0-\mu_0)\langle n\rangle - \frac{r_0}{K}\langle n^2\rangle$

$$\langle \dot{n^2}\rangle = 2\left(r_0-\mu_0 + \frac{r_0}{K}\right)\langle n^2\rangle + (r_0-\mu_0)\langle n\rangle \boxed{- \frac{3r_0}{K}\langle n\rangle} + \frac{2r_0}{K}\langle n^3\rangle$$

All models yield the same mean, but yield different higher moments

All models are well-mixed, birth & death models require fast equilibration of resource,
annihilation models require fast spatial equilibration of population, i.e. <u>not</u> diffusion-limited

Mass-action approximation $\langle n^2\rangle = \langle n\rangle\langle n\rangle$: $\boxed{\frac{dn}{dt} \simeq (r_0-\mu_0)\, n - \frac{r_0}{K} n^2}$

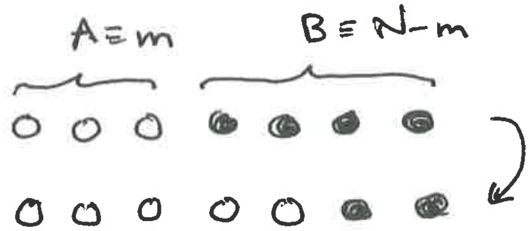
Other applications of discrete Master eqn.

* Wright-Fisher / Moran model
* Gene expression
* nucleation, self-assembly, biophysics, molecular motors

Wright-Fisher Model   1922, 1931

Consider a population of fixed size $N = n_A + n_B$ and of two species A & B.

discrete generations :



randomly select
an individual for replacement
N times in one "generation"

prob. of selecting A $= \frac{m}{N}$ ($\frac{3}{7}$ in this example)

after N selections, what is the probability of having n A types?

$$\boxed{P_{m,n} = \binom{N}{n} \left(\frac{m}{N}\right)^n \left(1 - \frac{m}{N}\right)^{N-n}}$$ one generation

state space for n_A $\{0, 1, 2, \ldots N\}$, 0 and N are both "absorbing" states

if no A particles, then extinction

* Expectation of n_A in $i+1$ generation given $n_A(i) = m$:

$$\mathbb{E}\left[n_A(i+1) \mid n_A(i)\right] = \sum_{n=0}^{N} n \binom{N}{n} \left(\frac{m}{N}\right)^n \left(1-\frac{m}{N}\right)^{N-n} = m$$

$\mathbb{E}[n_A(i)] = \mathbb{E}[n_A(0)] \Rightarrow$ expected n_A does not change

"neutral model"

$$\mathbb{V}\left[n_A(i+1) \mid n_A(i)\right] \equiv \mathbb{E}[n^2] - (\mathbb{E}[n])^2$$

$$= N\left(\frac{m}{N}\right)\left(1-\frac{m}{N}\right)$$

expected fraction of A is constant, but fluctuations lead to extinction ($n_A = 0$) or fixation ($n_A = N$) at the absorbing states

$\mathbb{E}[n_A] = \text{constant} = m$ (initial population of n_A)

at long times, mean n_A is $N\,P(n_A = N) + 0\,P(n_A = 0)$

$$\equiv \mathbb{E}[n_A(\infty)] = m$$

$$\boxed{\therefore\ P(n_A = N) = \frac{m}{N} = \text{initial fraction of } n_A}$$ "gambler's ruin"

probability of fixation

* Mean time to fixation (either boundary)

prob of going to n

$$\overline{T}_m = 1 + \sum_{n=0}^{N} p_{m,n} \overline{T}_n$$

mean time to extinction from n

one generation out of m (can return to m)

Continuum approximation:

$$\overline{T}_m \Rightarrow \overline{T}(x) \simeq 1 + \int_\Delta p(x \to x+\Delta)\, \overbrace{\overline{T}(x+\Delta)}^{\overline{T}(x) + \Delta \overline{T}'(x) + \frac{\Delta^2}{2}\overline{T}'' + \dots}$$

local jumps only

Δ is "small"

$$\approx 1 + \underbrace{\overline{T}(x)} + \overline{T}'(x) \int_\Delta p(x \to x+\Delta)\, \Delta$$

$$\overline{T}(x) \int_\Delta p(x \to x+\Delta) \equiv \overline{T}(x) \qquad + \tfrac{1}{2}\overline{T}''(x) \int_\Delta p(x \to x+\Delta)\, \Delta^2$$

$$\overline{T}(x) \approx 1 + \overline{T}(x) + \overline{T}'(x) \underbrace{\int_\Delta \Delta\, p(x \to x+\Delta)}_{\mathbb{E}(\Delta)} + \tfrac{1}{2}\overline{T}''(x) \int_\Delta p(x \to x+\Delta)\, \Delta^2$$

$$\bar{T}(x) \simeq 1 + \bar{T}(x) + \bar{T}'(x)\,\underbrace{\mathbb{E}(\Delta)} + \tfrac{1}{2}\bar{T}''(x)\,\underbrace{\mathbb{E}(\Delta^2)}$$

$$\mathbb{E}(X_{t+1} - X_t) = 0 \qquad \mathbb{E}(\Delta^2) = \mathbb{V}(\Delta) = \frac{X(1-X)}{N}$$

$$\therefore \quad \cancel{\bar{T}(x)} = 1 + \cancel{\bar{T}(x)} + \frac{1}{2N}(x(1-x))\,\bar{T}''(x)$$

$$\bar{T}''(x) = -\frac{2N}{X(1-X)},$$ integrating and using the boundary conditions $T(0) = T(1) = 0$,

$$\boxed{\bar{T}(x) \simeq -2N\left[x \ln x + (1-x)\ln(1-x)\right]}$$

finite $x \sim \mathcal{O}(1)$, $\bar{T} \sim N$

"diffusion"

These results are for "neutral model" Genetic "drift" only

— Now, consider selection s (this is a "convection")

if species A has an extra probability $1+s$ of being chosen to be a parent,

$$P_{m,n} = \binom{N}{n}\eta_m^n (1-\eta_m)^{N-n} \quad \text{where } \eta_m = \frac{(1+s)m}{(1+s)m + N - m} = \frac{m}{N} \text{ only for } s=0$$

expectations $\mathbb{E}[A]$, $\mathbb{E}[A^2]$ are simply found by replacing

$$\frac{m}{N} \rightarrow \frac{\eta}{N}$$

— Fixation probability

$$\boxed{P_{fix}(m) \cong \frac{1-e^{-ms}}{1-e^{-Ns}}}$$

starting population

$$\bar{T}(x = 1/N) = \frac{2N}{1+\frac{3}{4}Ns}\left(\frac{1+\frac{s}{2}\ln\frac{N}{2}}{1+s}\right) \quad \text{(conditioned on fixation)}$$

- Including mutations If A mutates to B with prob u, and B mutates to A with prob v, (at each generation),

$$\boxed{\begin{aligned} P_{mn} &= \binom{N}{n} \xi_m^n (1-\xi_m)^{N-n} \quad \text{where} \\ \xi_m &= (1-u)\eta_m + v(1-\eta_m) \end{aligned}}$$

These Wright-Fisher results are derived assuming non-overlapping generations

Moran Model

- At each time, one individual is randomly chosen to reproduce, and another is chosen to die.
- One parent and one offspring generated at each time point
- at each time, the species numbers can only increase or decrease by one or does not change

For neutral case;

$$P_{m,m-1} = \frac{m}{N} \frac{N-m}{N}$$

prob B was selected to reproduce

prob A was selected to die

$$P_{m,m+1} = \frac{N-m}{N} \frac{m}{N}$$

$$P_{m,m} = 1 - P_{m,m+1} - P_{m,m-1} = \frac{m^2 + (N-m)^2}{N^2}$$

Mean absorbing time

$$\boxed{\bar{T}(m) = N(N-1)\sum_{j=1}^{m}\frac{1}{N-j} + 2Nm\sum_{j=m+1}^{N-1}\frac{1}{j}}$$

Mean conditional fixation time

Conditioned on A fixating

$$\bar{T}(m|A) = \frac{N(N-m)}{m}\sum_{j=1}^{m}\frac{j}{N-j} + N(N-m-1) \sim N^2$$

$$\bar{T}(1|A) = N(N-1) \sim N^2$$

Compare with Wright-Fisher fixation time $\bar{T} \sim N$

extra factor of N results from "slower" timescale because only one parent selected at each time.

Continuous-time limit $\to$ Master equation

$$p_{n,n+1} \to r \cdot \frac{N-n}{N} \cdot \frac{n}{N}\, dt \equiv p_{n,n+1}\, dt \,, \quad p_{nn}(dt) = 1 - p_{n,n-1}dt - p_{n,n+1}dt$$

Prob of n A-species $P(n,t)$

$$\boxed{\to \frac{dP(n,t)}{dt} = p_{n-1,n} P(n-1,t) + p_{n+1,n} P(n+1,t) - (p_{n,n-1} + p_{n,n+1}) P(n,t)}$$

Taking the $N \to \infty$ limit and $\frac{n}{N} \equiv x$,

we can Taylor expand Master eqn:

$$\boxed{\frac{\partial P(x,t)}{\partial t} = \frac{r}{N^2} \frac{\partial^2}{\partial x^2}\left[x(1-x)P\right]}$$ forward eqn

Convection terms arise when species can mutate and under selection.

* Mean time to extinction or fixation

$\bar{T}_m \equiv$ mean "time" (number of generations) to hit $n_A = 0$ or $n_A = N$

$$= 1 + \sum_{n=0}^{N} p_{m,n} \bar{T}_n \qquad \text{(recursion relation)}$$

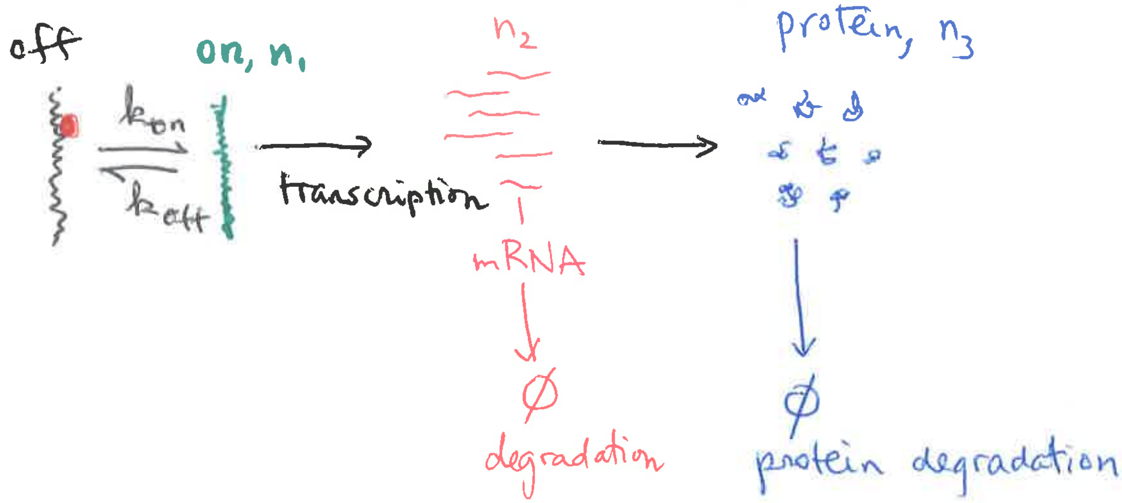
$\uparrow$ one generation to state n $\qquad \bar{T}_0 = \bar{T}_N = 0$

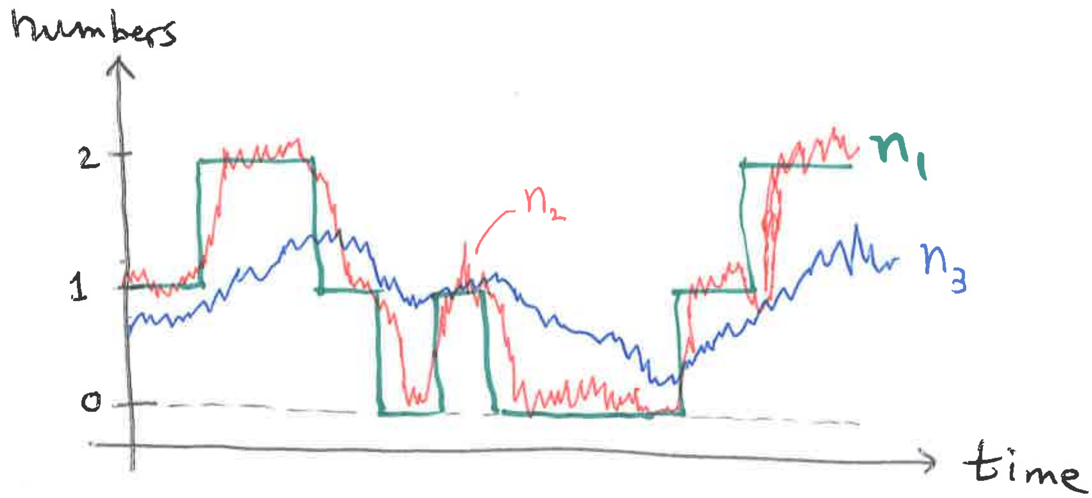
$m = 1, 2, \ldots N-1 \Rightarrow N-1$ coupled *linear* equations for $\bar{T}_m \neq 0$

Continuum approximation $\bar{T}(x) \simeq \bar{T}_m \simeq 1 + \int_{\Delta} p(x \to x+\Delta)\, \bar{T}(x+\Delta)$

$$\bar{T}(x+\Delta) \simeq \bar{T}(x) + \Delta \frac{\partial \bar{T}}{\partial x} + \frac{\Delta^2}{2} \frac{\partial^2 \bar{T}}{\partial x^2} + \cdots$$

Gene expression models





activation $\quad n_1 \xrightarrow{k_{on}(n_1^{max}-n_1)} n_1+1$

inactivation $\quad n_1 \xrightarrow{k_{off} n_1} n_1-1$

$$\frac{d\langle n_1\rangle}{dt} = k_{on}\left(n_1^{max} - \langle n_1\rangle\right) - k_{off}\langle n_1\rangle$$

transcription $\quad n_2 \xrightarrow{k_2} n_2+1$

mRNA degradation $\quad n_2 \xrightarrow{\mu_2} n_2-1$

$$\frac{d\langle n_2\rangle}{dt} = k_2\langle n_1\rangle - \mu_2\langle n_2\rangle$$

translation $\quad n_3 \xrightarrow{k_3} n_3+1$

protein degradation $\quad n_3 \xrightarrow{\mu_3} n_3-1$

$$\frac{d\langle n_3\rangle}{dt} = k_3\langle n_2\rangle - \mu_3\langle n_3\rangle$$

describe Master eqn for all 3 populations within a cell

on-off dynamics of DNA

$$\frac{dP(n_1, n_2, n_3, t)}{dt} = k_{on}(n_1^{max} - n_1 + 1)\,P(n_1 - 1, n_2, n_3) - k_{on}(n_1^{max} - n_1)\,P(n_1, n_2, n_3)$$

$$+ k_{off}(n_1 + 1)\,P(n_1 + 1, n_2, n_3) - k_{off}\, n_1\, P(n_1, n_2, n_3)$$

mRNA dynamics

$$+ k_2 \underline{n_1}\, P(n_1, n_2 - 1, n_3) - k_2 n_1 P(n_1, n_2, n_3)$$

$$+ \mu_2 (n_2 + 1)\, P(n_1, n_2 + 1, n_3) - \mu_2 n_2 P(n_1, n_2, n_3)$$

protein dynamics

$$+ k_3\, n_2\, P(n_1, n_2, n_3 - 1) - k_3 n_3 P(n_1\, n_2\, n_3)$$

$$+ \mu_3 (n_3 + 1)\, P(n_1, n_2, n_3 + 1) - \mu_3 n_3 P(n_1\, n_2\, n_3)$$

$\Rightarrow$ can find steady-state distribution $P_{ss}(n_1, n_2, n_3\; t \to \infty)$

(not equilibrium)

eigenvector corresponding to zero-eigenvalue, from

$P_{ss}(n_1, n_2, n_3)$, calculate moments.

$$CV_3^2 = \frac{\sigma_3^2}{\langle n_3\rangle^2} = \underbrace{\frac{1}{\langle n_3\rangle}}_{\text{birth-death of proteins}} + \overbrace{\frac{1}{\langle n_2\rangle}}^{\text{mRNA noise}} \frac{\mu_3}{\mu_2 + \mu_3} + \underbrace{\frac{1 - P_{on}}{\langle n_1\rangle} \cdot \frac{\mu_3}{\mu_2 + \mu_3} \cdot \frac{\mu_3}{k_{on} + k_{off} + \mu_3} \cdot \frac{\mu_2(k_{on} + k_{off} + \mu_2 + \ldots}{\mu_3(k_{on} + k_{off} + \mu_{\ldots}}}_{\text{DNA on-off}}$$

(DNA on)

$$CV_2^2 = \frac{\sigma_2^2}{\langle n_2 \rangle^2} = \frac{1}{\langle n_2 \rangle} + \frac{1 - P_{on}}{\langle n_1 \rangle} \frac{\mu_2}{\mu_2 + k_{on} + k_{off}}$$

$$CV_1^2 = \frac{\sigma_1^2}{\langle n_1 \rangle^2} = \frac{1 - P_{on}}{\langle n_1 \rangle}$$

"Fano Factor" $\langle n \rangle CV^2$

$$\Rightarrow \frac{\sigma_3^2}{\langle n_3 \rangle} \simeq 1 + \frac{\langle n_3 \rangle}{\langle n_2 \rangle} \frac{\mu_3}{\mu_2 + \mu_3} \quad \text{for } \underbrace{\mu_3 << \mu_2}_{\text{proteins last longer than their mRNAs}}$$

> 1 (noise greater than that of simple Poisson process)

"bursts" of proteins?

not necessarily, need to look at trajectories $n_3(t)$ to look for "bursts" of protein production.

* one mRNA can produce many proteins before degrading.

Steady-state distributions

consider on and off states and protein (neglect mRNA intermediate)

J. E. Hornos, D. Schultz, G. Innocentini, Jin. Wang, A. M. Walczak, J. Onuchic, P. G. Wolynes:

PRE, 72 2005

assume two states of DNA with bath of n proteins

$\alpha_n \equiv$ prob of unbound DNA (active)

$\beta_n \equiv$ prob of protein-bound DNA (inactive)

protein production rate of active DNA

binding (inactivation) rate

$$\frac{d\alpha_n}{dt} = g_\alpha(\alpha_{n-1}-\alpha_n) + \mu[(n+1)\alpha_{n+1} - n\alpha_n] - h n\alpha_n + f\beta_n \boxed{+\mu_b\beta_{n+1}}$$

degradation rate

degradation of bound protein

unbinding (activation) rate

$$\frac{d\beta_n}{dt} = g_\beta(\beta_{n-1}-\beta_n) + \mu[(n+1)\beta_{n+1} - n\beta_n] + h n\alpha_n - f\beta_n \boxed{-\mu_b\beta_n}$$

$(n \geq 2)$

protein production rate of inactive DNA

$(g_\beta \ll g_\alpha)$

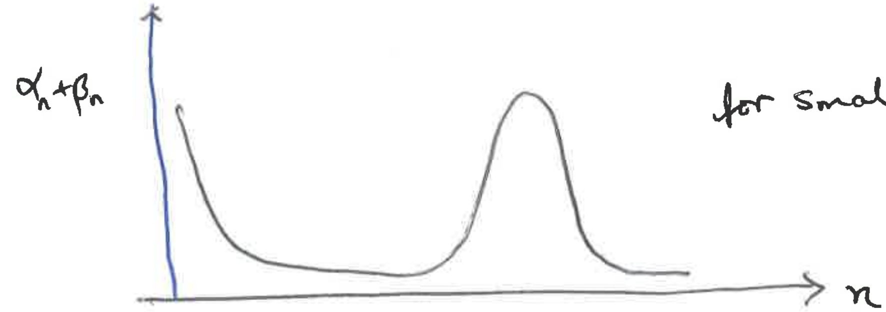
+ Boundary conditions, e.g. $\frac{d\alpha_0}{dt} = -g_\alpha\alpha_0 + \mu(\alpha_1+\beta_1)$, etc

solved using generating functions $\alpha(z) \equiv \sum_{n=0}^{\infty}\alpha_n z^n$, $\beta(z) = \sum_{n=0}^{\infty}\beta_n z^n$

$$\Longrightarrow \frac{d\alpha(z)}{dz} = \ldots \quad \frac{d\beta(z)}{dz} \xrightarrow[\text{combine}]{} \frac{d^2\alpha(z)}{dz^2} + p\frac{d\alpha(z)}{dz} + q\alpha(z) = 0$$

$\Rightarrow \alpha_n \sim$ Kummer functions

conclusion: double-peaked steady-state distribution $\alpha_n + \beta_n$



for small f/μ

this result is important: heterogeneity in protein populations

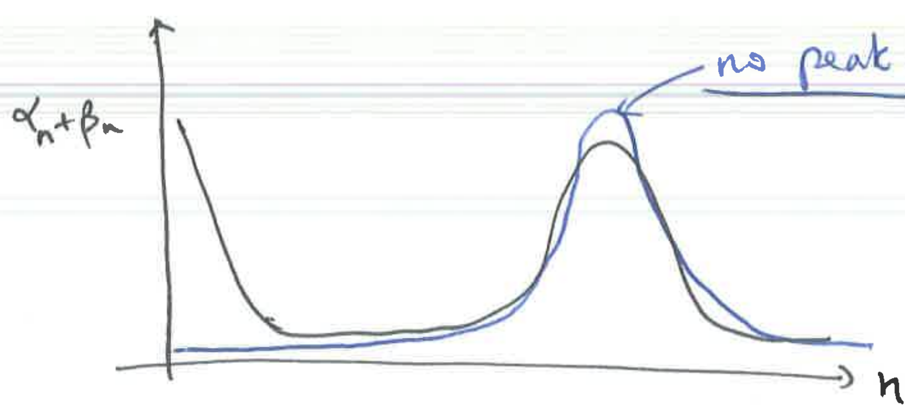
---

However, the original $\boxed{\text{equations}}$ are not mechanistically consistent: proteins also degrade while bound — as assumed in Hornos et-al, but equations were wrong.

However, maybe $\mu_b = 0$? Then, equations look like those of Hornos. <u>But</u> then boundary condition

$$\frac{d\alpha_0}{dt} = -g_\alpha \alpha_0 + \mu(\alpha_1 + \beta_1)$$ is <u>wrong</u>!

using same parameters with the correct model, Grima, Schmidt, & Newman, JCP, <u>137</u> 2012 show



no peak!

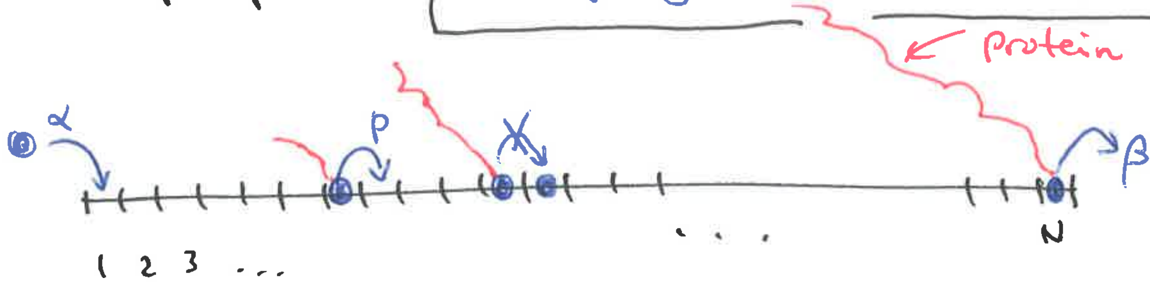
$\boxed{\text{7 years to correct!}}$

* <u>Give Slides on nucleation here</u>

Look more closely at mRNA

mRNA's produce proteins using ribosomes.

Simple picture: $\boxed{\text{Totally asymmetric exclusion process (TASEP)}}$



- ribosomes only move forward when a proper amino acid binds and is added to the protein

- first, assume p is constant

  if α is large (β-small) — high density

  if β is large (α-small) — low density

- what is steady state current?

Stochastic process (Markov) for occupation configurations on a single mRNA strand: 2^N states,

transition matrix $2^N \times 2^N$

- Exact solution 1992, 1993 Derrida, Schutz, Evans, Hakim, Pasquier

  → thousands of papers applying this model to many different biophysics problems.

Outline of solution in steady-state

$$\frac{\partial P(\sigma_1, \sigma_2, \dots \sigma_N; t)}{\partial t} = \text{in} - \text{out} = 0 \qquad 2^N \times 2^N$$

find zero eigenvalue & eigenvector
$$\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{2^N} \end{pmatrix}$$

Current $J = \beta P(\sigma_N = 1)$

$= \alpha P(\sigma_1 = 0) = \dots$

alternative approach — Derrida, Evans, Hakim, Pasquier, J Phys A, Math. 26 1493-1517, (1993)

$$P_N(\sigma_1, \sigma_2, \dots \sigma_N) \equiv \frac{f_N(\sigma_1, \sigma_2, \sigma_3 \dots \sigma_N)}{Z_N}$$

$$Z_N = \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} \dots \sum_{\sigma_N=0,1} f_N(\sigma_1, \sigma_2, \dots \sigma_N)$$

"Guess" an algebra
$$f_N(\sigma_1, \sigma_2, \dots \sigma_N) = \langle W| \prod_{i=1}^{N} [\sigma_i D + (1-\sigma_i) E] |V\rangle$$

right vector

left vector

E and D are matrices s.t.

$$DE = D + E \equiv C, \qquad D|V\rangle = \frac{1}{\beta}|V\rangle, \qquad \langle W|E = \frac{1}{\alpha}\langle W|,$$

$$\langle W|V\rangle = 1$$

$$Z_N = \langle W|C^N|V\rangle, \qquad \langle \sigma_i \rangle = \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} \dots \sum_{\sigma_N=0,1} \sigma_i \frac{f_N(\sigma_1, \sigma_2 \dots \sigma_N)}{Z_N}$$

$$= \frac{\langle W|C^{i-1} D C^{N-i}|V\rangle}{\langle W|C^N|V\rangle}$$

Conservation of current

$$\alpha \langle W | E C^{N-1} | V \rangle = \ldots = \langle W | C^{i-1} D E C^{N-i-1} | V \rangle$$

$$= \ldots = \beta \langle W | C^{N-1} D | V \rangle$$

using algebra,

$$C^n = \sum_{k=0}^{n} \frac{k(2n-1-k)!}{n!\,(n-k)!} \sum_{q=0}^{k} E^q D^{k-q}$$, ~~Alcursing~~,

$$\langle W | C^n | V \rangle = \sum_{k=0}^{n} \frac{k(2n-1-k)!}{n!\,(n-k)!} \sum_{q=0}^{k} \frac{1}{\alpha^q} \frac{1}{\beta^{k-q}}$$

$$= \sum_{k=0}^{n} \frac{k(2n-1-k)!}{n!\,(n-k)!} \frac{(1/\beta)^{k+1} - (1/\alpha)^{k+1}}{(1/\beta) - (1/\alpha)}$$

$$\boxed{J = \frac{\langle W | C^{N-1} | V \rangle}{\langle W | C^N | V \rangle} = \frac{S_{N-1}(1/\beta) - S_{N-1}(1/\alpha)}{S_N(1/\beta) - S_N(1/\alpha)}}$$

$$\boxed{S_N(x) = \sum_{k=0}^{N-1} \frac{(N-k)\,(N-1+k)!}{N!\,k!} x^{N-k+1}}$$

Jm

Steady-state current can be found exactly for $N \gg 1$ by assuming mean-field dynamics

occupation at site i: σ_i

mean field approx: $\langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle$

Conservation of steady-state current:

$$\langle \alpha (1-\sigma_1) \rangle = J$$

$$\langle p\, \sigma_i (1-\sigma_{i+1}) \rangle = J \quad \forall\, i$$

$$\langle \beta \sigma_N \rangle = J$$

if $p \langle \sigma_i (1-\sigma_{i+1}) \rangle \cong p \langle \sigma \rangle (1-\langle \sigma \rangle)$

$= \alpha (1-\langle \sigma \rangle)$, $\sigma_1 \sim \sigma_i$

- $\langle \sigma \rangle = \dfrac{\alpha}{p}$, $J = \alpha \left(1-\dfrac{\alpha}{p}\right)$

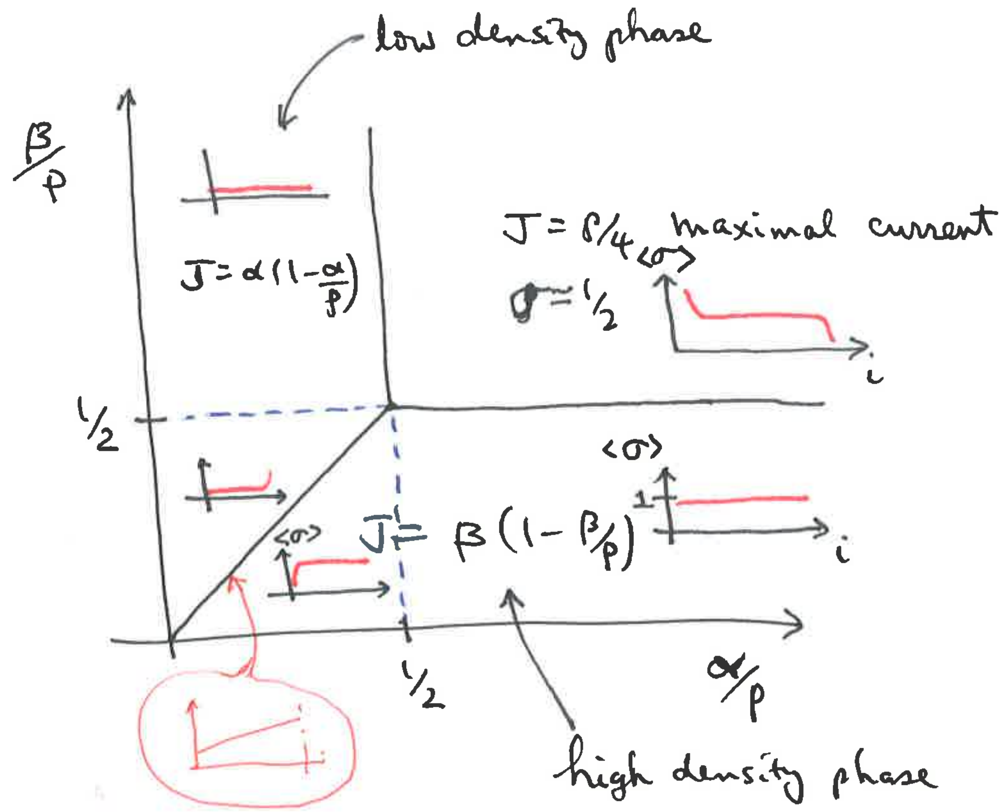
if $p \langle \sigma \rangle (1-\langle \sigma \rangle) = \beta \langle \sigma \rangle$, $\sigma_N \sim \sigma_i$

- $\dfrac{\beta}{p} = 1-\langle \sigma \rangle$, $J = \beta \left(1-\dfrac{\beta}{p}\right)$

if σ_1, σ_N not directly equal to σ_i,

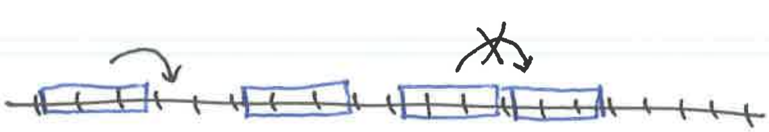
$p \langle \sigma \rangle (1-\langle \sigma \rangle) = J \Rightarrow$ maximize wrt $\langle \sigma \rangle$

- $\Rightarrow \langle \sigma \rangle = 1/2$, $J = \dfrac{p}{4}$

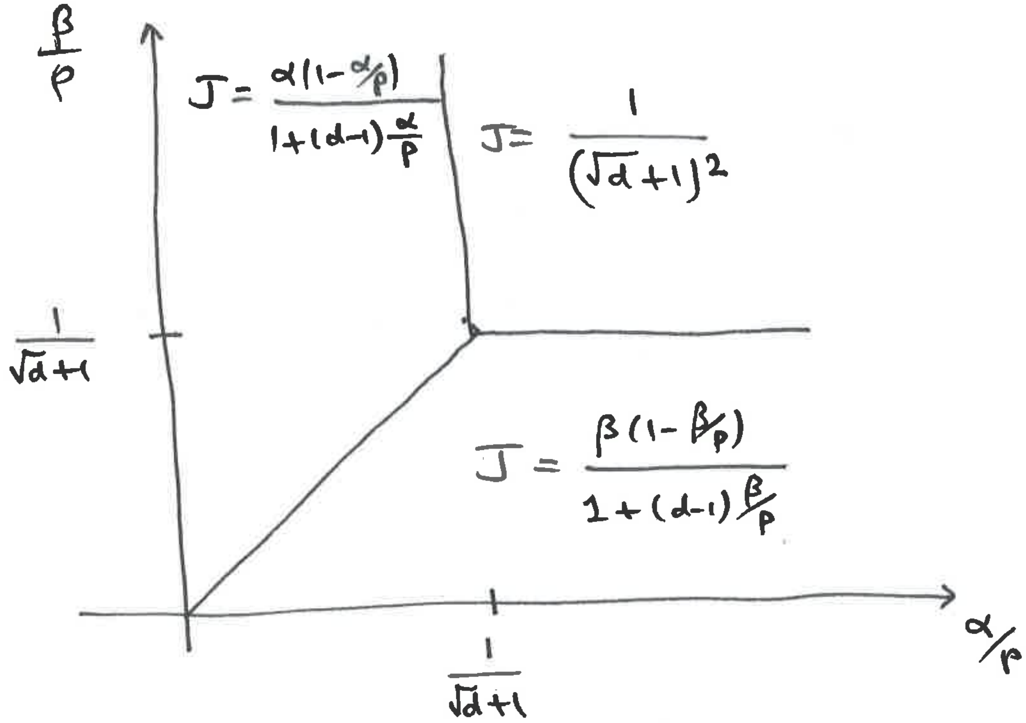


<u>Many extensions have been developed:</u>

- heterogeneous internal hopping rates (sequence dependence) p_i
- partially asymmetric models (backward hopping, rate q)
- fluctuations
- multiple species
- multiple parallel lattices with particles hopping between lattices
- variable lattice sizes N
- Langmuir kinetics (particles hopping on and off the lattice)
- particles of arbitrary size, each particle (ribosome) occupies $\ell > 1$ site.



$\ell = 3$



If density is smoothly varying, we can develop hydrodynamic equations for the mean occupation $\sigma(x,t)$ where $x = i/N$:

$$\frac{\partial \sigma(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(p(x)\,\sigma\, G(\sigma)\right) + \frac{1}{N}\frac{\partial^2}{\partial x^2}\left(p(x)\, G(\sigma)\right)$$

where $G(\sigma) = \dfrac{1-\ell\sigma}{1-(\ell-1)\sigma}$ and $p(x)$ is the local hopping rate

+ boundary conditions at $x=0$, $x=1$