ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Diffusion theory

Niko Beerenwinkel

Outline

- Kolmogorov forward equation

- Kolmogorov backward equation

- Fixation probability

- Mean fixation time

Diffusion theory

- Intuitively, selection should dominate large populations, while drift should dominate small populations.

- Diffusion theory combines all evolutionary forces in a single model that is valid for all population sizes:
  - Mutation
  - Selection
  - Drift
  - Recombination
  - Migration

- In particular, diffusion theory considers both deterministic and stochastic processes.

The structure of diffusion theory

- We follow the probability distribution (rather than the deterministic trajectory) of a population over time.

- The probability density that an allele A_1 is at frequency p at time t is
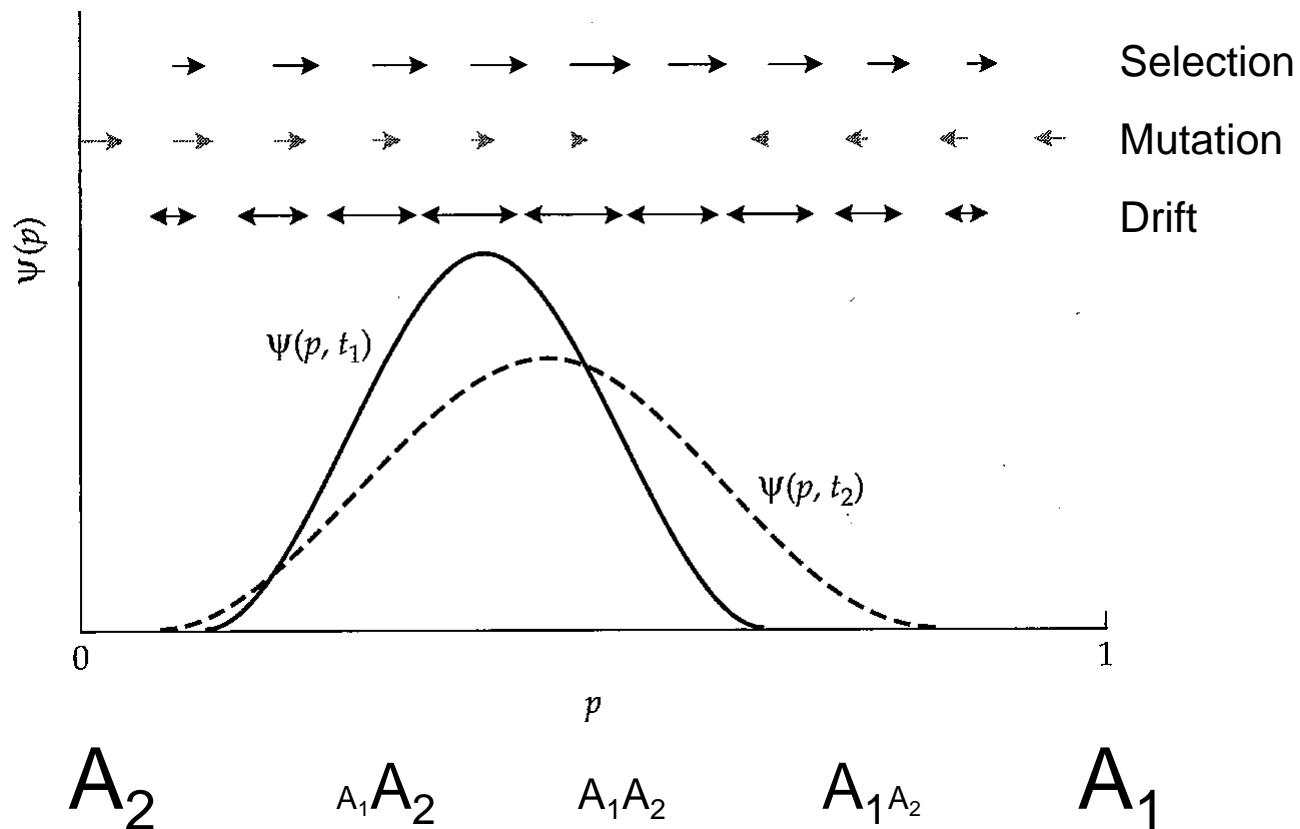
$$\psi(p, t)$$

- The allele frequency distribution $\psi(p) = \psi(p, \cdot)$ changes over time.

Two classes of evolutionary forces

- Directional processes, M(p)
  - Non-zero expected change in allele frequency
  - E.g., selection, mutation, migration, recombination
  - M(p) is measured as the expected change per generation

- Nondirectional processes, V(p)
  - Zero expected change
  - Increase in the spread of the allele distribution
  - E.g., random variation in survivorship, random gamete success
  - V(p) is measured by the expected variance in the next generation

Directional and non-directional processes

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Example: Moran process

- With p = p(t) = X(t)/N,

$$M(p) \ = \ E[p(t+1) - p(t) \mid p(t)]$$
$$= \ p - p = 0$$

$$V(p) \ = \ E\{\mathsf{Var}[p(t+1)] \mid p(t)\}$$
$$= \ \frac{1}{N^2} E\{\mathsf{Var}[X(t+1)] \mid X(t)\}$$
$$= \ \frac{2p(1-p)}{N^2}$$

Change in allele frequency

- We want to predict the change in allele frequency distribution from the current distribution.

- $\psi \quad = \quad \psi(p, t)$
  $\quad = \quad$ Probability density of allele frequency p at time t

- $g \quad = \quad g(p, \epsilon; dt)$
  $\quad = \quad$ Probability that the allele frequency changes from p to $p + \epsilon$ in time interval dt.

- Any allele frequency p at time t + dt must have evolved from some distribution p − ϵ at time t,

$$\psi(p, t + dt) = \int \psi(p - \epsilon, t) g(p - \epsilon, \epsilon; dt) d\epsilon$$

Taylor expansion

- Taylor series expansion of $\psi \cdot g$ around p:

$$\psi(p, t + dt) = \int \left[\psi(p,t) g(p, \epsilon; dt) - \right.$$

$$\left. - \epsilon \frac{\partial(\psi g)}{\partial p} + \frac{\epsilon^2}{2} \frac{\partial^2(\psi g)}{\partial p^2} - \frac{\epsilon^3}{6} \frac{\partial^3(\psi g)}{\partial p^3} + \ldots \right] d\epsilon$$

- The *diffusion approximation* is derived by assuming $\epsilon^2 \gg \epsilon^3$. This assumption means that the population does not change too much in any short time interval, i.e., the effects of selection and mutation are relatively weak.

Taylor approximation

- Also, p and ϵ are independent. Thus, approximately,

$$\psi(p, t + dt) = \psi(p, t) \int g(p, \epsilon; dt)\, d\epsilon -$$

$$- \frac{\partial}{\partial p}\psi \int g\epsilon\, d\epsilon + \frac{1}{2}\frac{\partial^2}{\partial p^2}\psi \int g\epsilon^2\, d\epsilon$$

- Because $\int g(p, \epsilon; dt)\, d\epsilon = 1$, we have

$$\psi(p, t + dt) = \psi - \frac{\partial}{\partial p}\psi \int g\epsilon\, d\epsilon + \frac{1}{2}\frac{\partial^2}{\partial p^2}\psi \int g\epsilon^2\, d\epsilon$$

What do the integrals mean?

$$\psi(p, t + dt) = \psi - \frac{\partial}{\partial p}\psi \int g\epsilon \, d\epsilon + \frac{1}{2}\frac{\partial^2}{\partial p^2}\psi \int g\epsilon^2 \, d\epsilon$$

- $\int g\epsilon \, d\epsilon = E(\epsilon)$ is the expected change over time dt. With M(p) the rate of directional change, we have

$$\int g(p, \epsilon; dt)\epsilon \, d\epsilon = M(p)dt$$

- $\int g\epsilon^2 \, d\epsilon = E(\epsilon^2) = Var(\epsilon) + E(\epsilon)^2 \approx Var(\epsilon)$. With V(p) the variance in allele frequency due to nondirectional effects,

$$\int g(p, \epsilon; dt)\epsilon^2 \, d\epsilon = V(p)dt$$

Integrals substituted

- Substituting we obtain

$$\psi(p, t + dt) = \psi(p, t) -$$
$$- \frac{\partial[\psi(p, t)M(p)]}{\partial p} dt + \frac{1}{2}\frac{\partial^2[\psi(p, t)V(p)]}{\partial p^2} dt$$

- Now substract ψ(p, t) and divide by dt.
- For dt $\rightarrow$ 0, we obtain ….

The Kolmogorov forward equation

$$\frac{\partial \psi(p,t)}{\partial t} = -\frac{\partial}{\partial p}\left[\psi(p,t)M(p)\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(p,t)V(p)\right]$$

- The equation describes the change in allele frequency in terms of the shape of that distribution and the functions M (directional process) and V (nondirectional process).
- It is also known as the *Fokker-Planck equation* or the *diffusion equation*.

Equilibrium

- At equilibrium,

$$-\frac{\partial}{\partial p}\left[\psi^*(p,t)M(p)\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi^*(p,t)V(p)\right] = 0$$

- Integrating over p yields

$$\frac{1}{2}\frac{\partial}{\partial p}\left[\psi^*(p,t)V(p)\right] - \psi^*(p,t)M(p) = 0$$

- This first order homogeneous ODE can be solved.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Equilibrium distribution

- The equilibrium distribution of the allele frequency is

$$\psi^*(p) \;=\; \frac{C}{V(p)} \exp\left[\int_0^p \frac{2M(q)}{V(q)} dq\right]$$

where C is a constant of integration.

Selective competition between two alleles

- Assume two alleles A_1 and A_2 with frequencies p and $1 - p$, and fitness w_1 and w_2, respectively.

- The average fitness of the population is

$$\bar{w} = pw_1 + (1 - p)w_2 \quad \text{and} \quad \frac{d\bar{w}}{dp} = w_1 - w_2$$

- In the next generation, the allele frequency is

$$p' = (pw_1)/\bar{w}$$

Wright's equation for an adaptive landscape

$$\triangle p_{\mathsf{sel}} = p' - p = \frac{p(w_1 - \bar{w})}{\bar{w}}$$

$$= \frac{p(1-p)(w_1 - w_2)}{\bar{w}}$$

$$= \frac{p(1-p)}{\bar{w}} \frac{d\bar{w}}{dp}$$

$$= p(1-p) \frac{d \log \bar{w}}{dp}$$

Mutation and selection

- Let u_1 be the A_1-to-A_2 mutation rate, and u_2 the A_2-to-A_1 mutation rate.

- The per-generation change due to mutation is
  $\triangle p_{mut} = -p\, u_1 + (1 - p)\, u_2$.

- Thus, the directional processes of selection and mutation add up to

$$M(p) = \frac{p(1-p)}{\bar{w}}\frac{d\bar{w}}{dp} - pu_1 + (1-p)u_2$$

Sampling variance

- Assuming Wright-Fisher dynamics, we have

$$
\begin{aligned}
V(p) &= E\{\mathrm{Var}[p(t+1)] \mid p(t)\} \\
&= E\{\mathrm{Var}[X(t+1)/N] \mid X(t)/N\} \\
&= \frac{1}{N^2} Np(1-p) \\
&= \frac{p(1-p)}{N}
\end{aligned}
$$

Mutation, selection, and sampling

$$\frac{M}{V} = N \left[\frac{d \log \bar{w}}{dp} - \frac{u_1}{1-p} + \frac{u_2}{p} \right]$$

$$\Rightarrow$$

$$\int_0^p \frac{M(q)}{V(q)} dq =$$

$$= N \left[\log \bar{w} + u_1 \log(1-p) + u_2 \log(p) \right]$$

Wright-Fisher dynamics with mutation and selection at equilibrium

$$\psi^*(p) = \frac{C}{V(p)} \exp\left[\int_0^p \frac{2M(q)}{V(q)}dq\right]$$

$$= C\,\bar{w}^{2N}(1-p)^{2Nu_1-1}p^{2Nu_2-1}$$

where C is given by the requirement that

$$\int_0^1 \psi^*(p)dp = 1$$

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

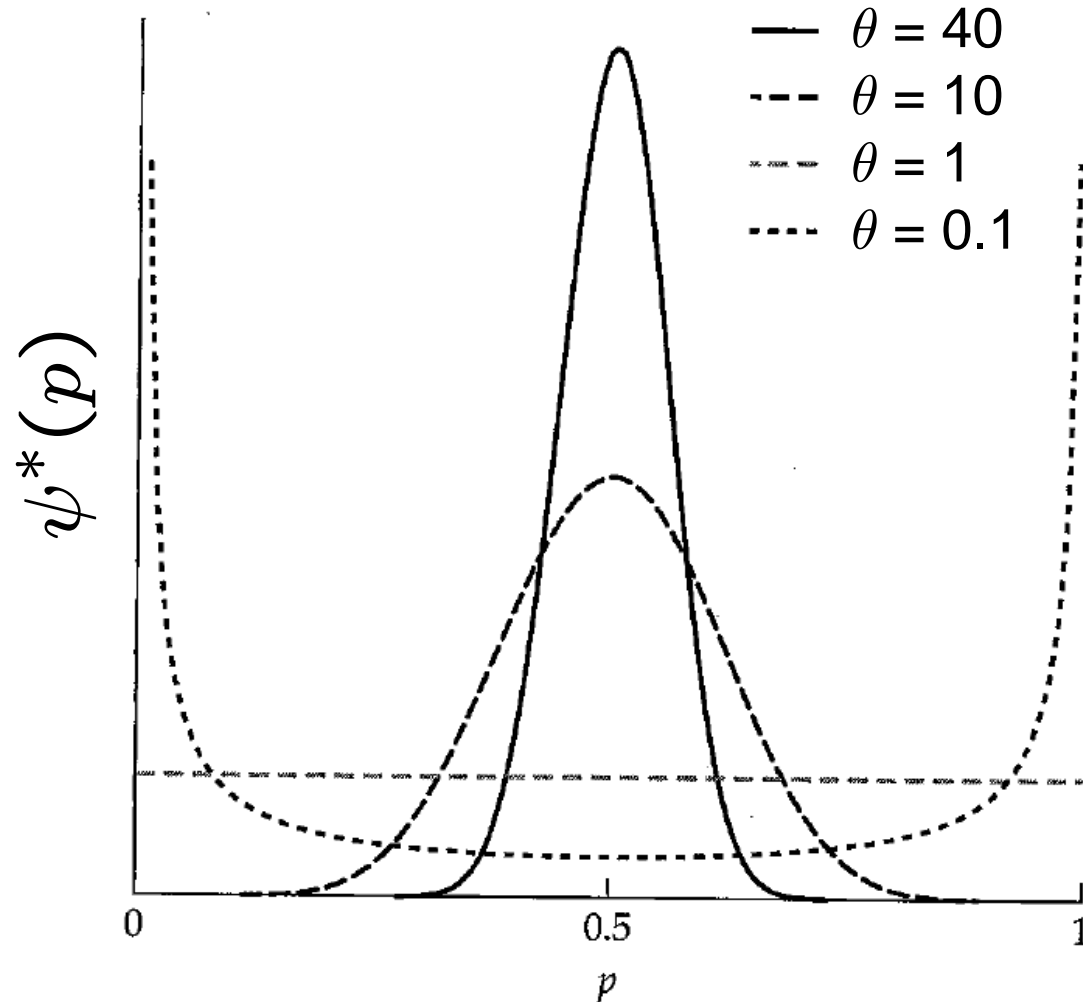D-BSSE
Department of Biosystems
Science and Engineering

How much neutral variation is maintained under mutation and drift?

For $\bar{w} = 1$ and $u = u_1 = u_2,$

$$\psi^*(p) \propto [p(1-p)]^{2Nu-1}$$

- We set $\theta = 2\,N\,u,$ because this scaled mutation parameter determines the equilibrium distribution.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich

D-BSSE
Department of Biosystems
Science and Engineering

Mutation-drift dynamics

Selection

- Assume that A_1 has fitness $w_1 = 1 + s$ and A_2 has fitness $w_2 = 1$. Then the average population fitness is
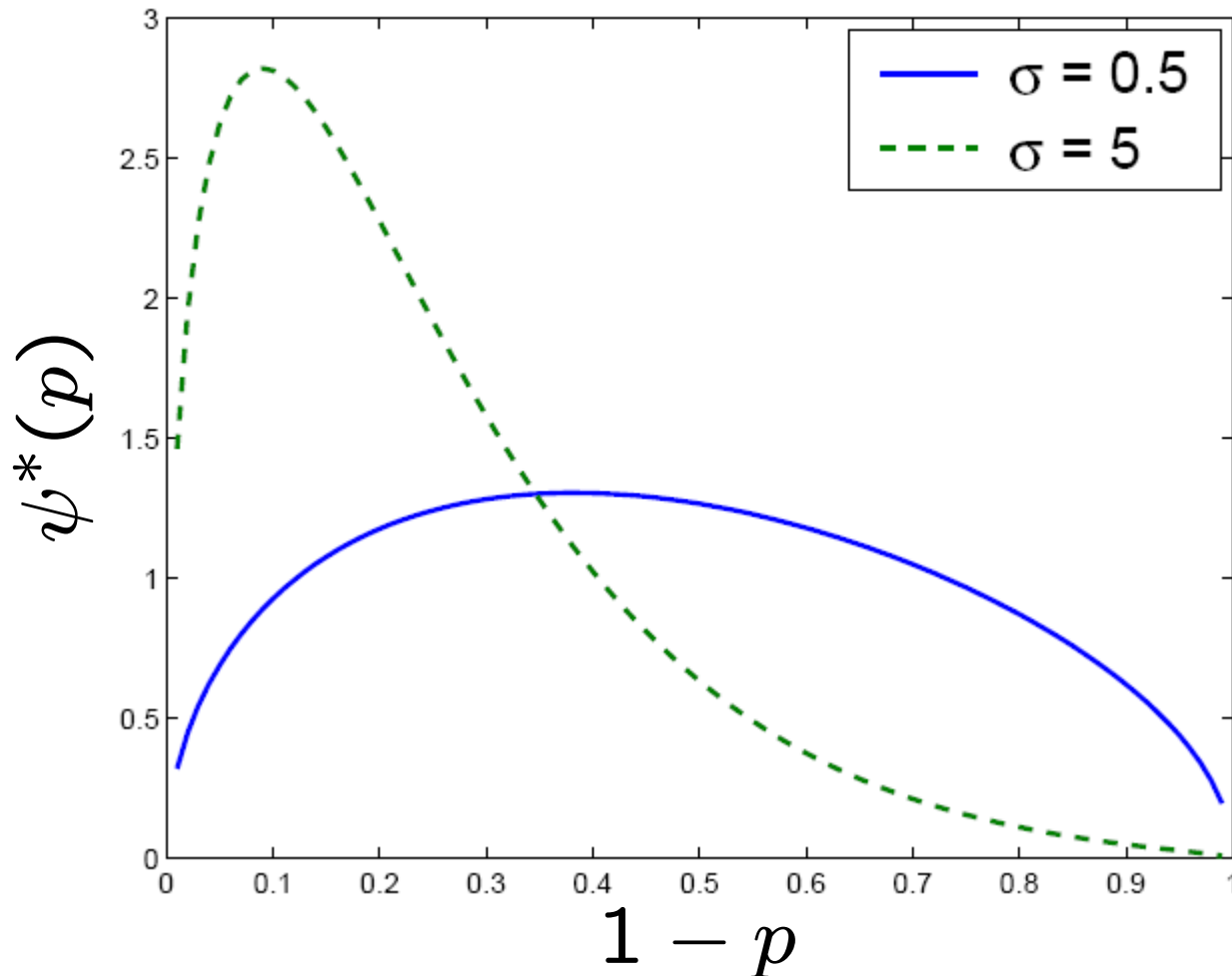
$$\bar{w} \; = \; p(1 + s) + (1 - p)$$
$$= \; 1 + sp \; \approx \; e^{sp}$$

$$\Rightarrow \; \bar{w}^{2N} \; = \; e^{2Nsp}$$

- We set $\sigma = 2\,N\,s$, because this scaled selection parameter determines the equilibrium dynamics.

- Ignoring mutation, we have

$$\psi^*(p) \propto e^{\sigma p}/p(1 - p)$$

Selection-drift dynamics

Summary

- The equilibrium allele frequency distribution in the diffusion approximation of the Wright-Fisher process (for haploids) reads

$$\psi^*(p) \propto (1-p)^{\theta-1}\, p^{\theta-1}\, e^{\sigma p}$$

with scaled mutation parameter θ = 2Nu and scaled selection parameter σ = 2Ns.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

What about fixation?

- We need to study ψ(p, t | p$_0$), the density of a population having allele frequency p at time t *given that the allele frequency was p$_0$ at time 0*.

- Consider all ways to go from p$_0$ to p$_0$ + ϵ in time dt:

$$\psi(p, t + dt \mid p_0) = \int \psi(p, t \mid p_0 + \epsilon) g(p_0, \epsilon; dt) d\epsilon$$

Taylor expansion

- Taylor series expansion of $\psi(p, t \mid p_0 + \epsilon)$ around p_0:

$$\psi(p, t \mid p_0 + \epsilon) = \psi(p, t \mid p_0) +$$

$$+ \frac{\partial \psi}{\partial p_0} + \frac{\epsilon^2}{2}\frac{\partial^2 \psi}{\partial p_0^2} + \frac{\epsilon^3}{6}\frac{\partial^3 \psi}{\partial p_0^3} + \ldots$$

- Write $\psi = \psi(p, t, \mid p_0)$, $g = g(p_0, \epsilon, dt)$ and approximate to

$$\psi(p, t + dt \mid p_0) = \psi \int g \, d\epsilon +$$

$$+ \frac{\partial \psi}{\partial p_0} \int g \epsilon \, d\epsilon + \frac{1}{2}\frac{\partial^2 \psi}{\partial p_0^2} \int g \epsilon^2 \, d\epsilon$$

The Kolmogorov backward equation

- The integrals have the same meaning as before.
- Subtract $\psi(p, t \mid p_0)$, divide by dt, and let dt $\rightarrow$ 0 to obtain

$$\frac{\partial \psi(p, t \mid p_0)}{\partial t} =$$

$$= M(p_0)\frac{\partial \psi(p, t \mid p_0)}{\partial p_0} + \frac{1}{2}V(p_0)\frac{\partial^2 \psi(p, t \mid p_0)}{\partial p_0^2}$$

Equilibrium distribution

- At equilibrium,

$$M(p_0)\frac{\partial \psi^*}{\partial p_0} + \frac{1}{2}V(p_0)\frac{\partial^2 \psi^*}{\partial p_0^2} = 0$$

- We find

$$\frac{\partial \psi^*}{\partial p_0} = C \exp\left(-\int_0^p \frac{2M(q)}{V(q)}dq\right)$$

- Integration w.r.t. p_0 does not yield a unique solution.

Probability of fixation

- We are interested in the probability of fixation of the first allele, A_1, ($p = 1$) given its initial allele frequency p_0, i.e., in

$$\rho(p_0) = \psi(1, \infty \mid p_0)$$

- Clearly,

$$\rho(1) = \psi(1, \infty \mid 1) = 1$$
$$\rho(0) = \psi(1, \infty \mid 0) = 0$$

- With these boundary conditions, we can solve for ψ^*.

Fixation probability

$$\rho(p_0) = \frac{\int_0^{p_0} \exp\left(-\int_0^p \frac{2M(q)}{V(q)} dq\right) dp}{\int_0^1 \exp\left(-\int_0^p \frac{2M(q)}{V(q)} dq\right) dp}$$

Selection

- Recall that

$$\bar{w} = 1 + ps, \qquad \frac{d\bar{w}}{dp} = s,$$

$$M(p) = \frac{p(1-p)}{\bar{w}}\frac{d\bar{w}}{dp} = \frac{p(1-p)s}{1+ps}$$

- For $p \ll 1$,

$$\frac{2M(p)}{V(p)} = \frac{2Ns}{1+ps} \approx 2Ns$$

because s is also small.

Fixation probability in a haploid population

- Now,

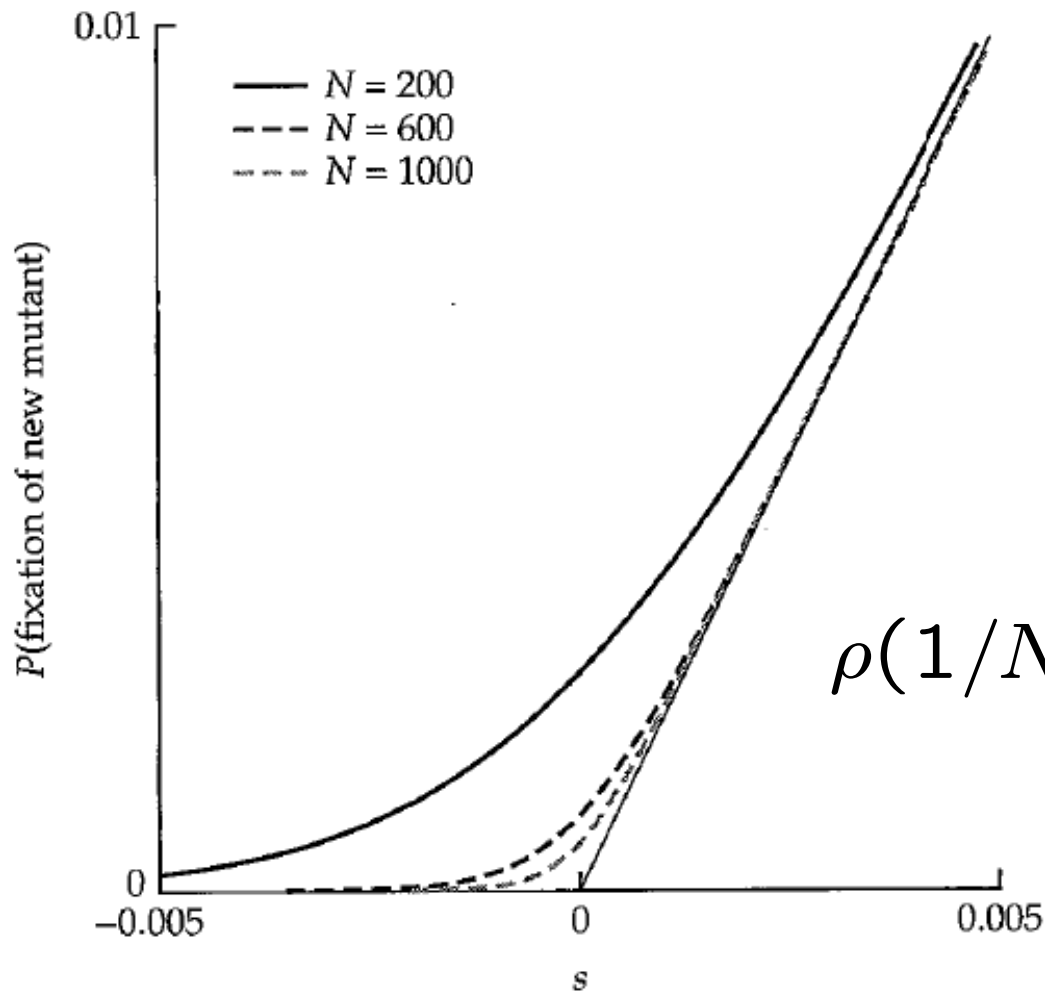$$\int_0^p \frac{2M(q)}{V(q)} dq = 2Nsp$$

and hence

$$\rho(p_0) = \frac{1 - e^{-2Nsp_0}}{1 - e^{-2Ns}}$$

Neutral case

- For s = 0, $\rho(p_0)$ is undefined. However,

$$\lim_{s \to 0} \rho(p_0) = \frac{-2Nsp_0}{-2Ns} = p_0$$

Fixation probability of a new mutant

$$\rho(1/N) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}$$

A new mutant in a large population

$$\rho(1/N) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}$$

- If N is large, $2Ns \gg 1$, then the denominator is one.
- If s is small, $s \ll 1$, then $\exp(-2s) \approx 1 - 2s$
- In this situation, we recover a classical result due to Haldane (1927):

$$\rho(\text{new mutant}) \approx 2s$$

Mean fixation time

- Let $\tau(p_0)$ be the expected waiting time until A_1 becomes fixed, given its fixation and the initial allele frequency p_0.

- Define

$$A(p) = \int_0^p \frac{2M(q)}{V(q)} dq$$

$$S(p_0) = \int_0^{p_0} \exp\left(-A(p)\right) dp$$
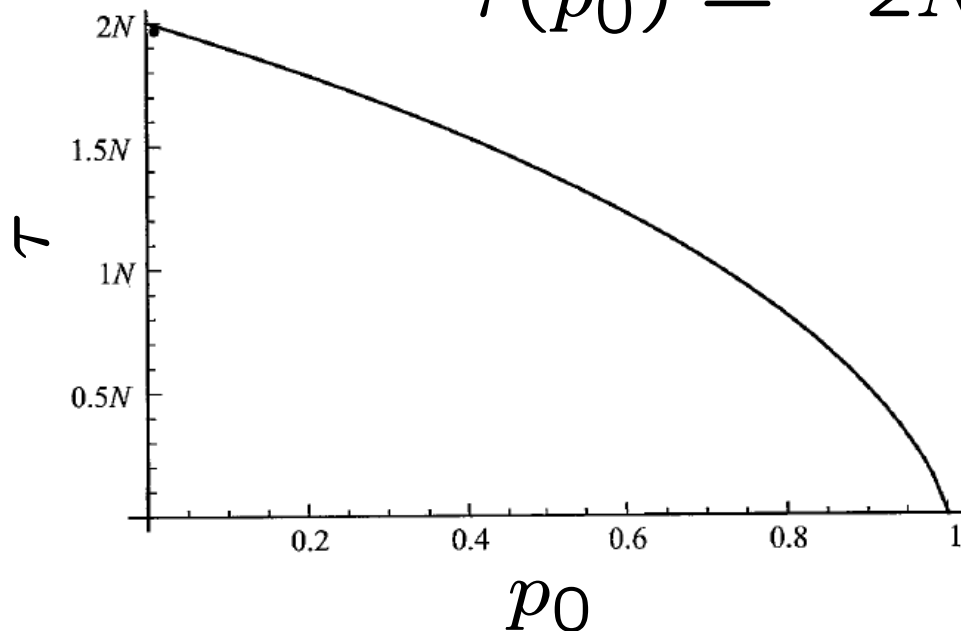
such that $S(0) = 0$ and $\rho(p_0) = S(p_0)/S(1)$

Waiting time to fixation

- Using the Kolmogorov backward equation, one can derive an analytical expression for $\tau(p_0)$.

$$\tau(p_0) = 2S(1)\left[\int_{p_0}^1 \frac{\rho(y)(1-\rho(y))}{\exp(-A(y))V(y)}dy+\right.$$

$$\left.+\frac{1-\rho(p_0)}{\rho(p_0)}\int_0^{p_0}\frac{\rho(y)^2}{\exp(-A(y))V(y)}dy\right]$$

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
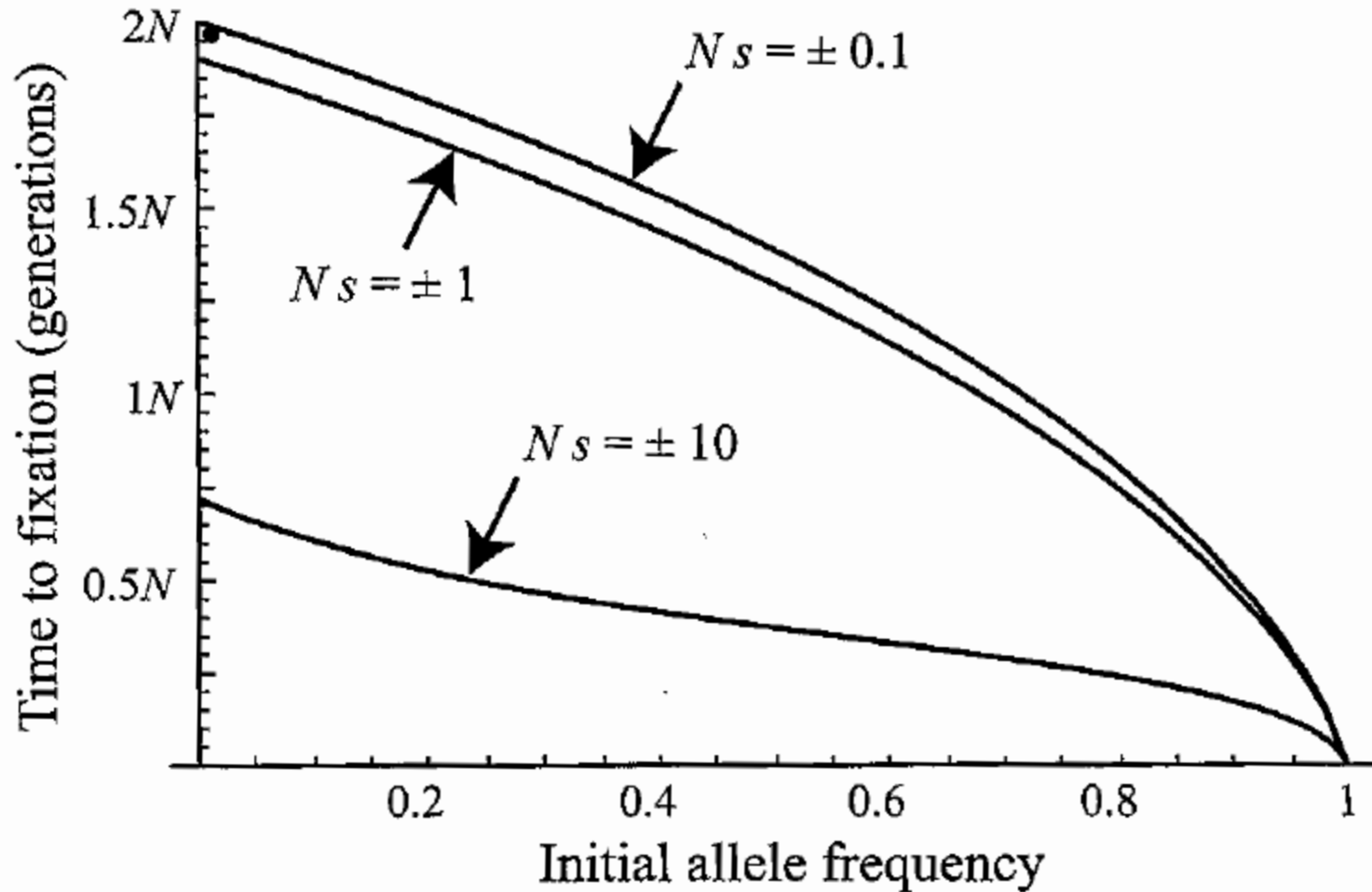Department of Biosystems
Science and Engineering

Fixation time for the neutral Wright-Fisher model

$$\tau(p_0) = -2N \frac{1 - p_0}{p_0} \log(1 - p_0)$$

- For $p_0 = 1 / N$, we have approximately $\tau(1/N) = 2N$

Waiting time for a selected mutant to fix

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Further reading

- Rice SH. *Evolutionary Theory*, Chapter 5

- Otto SP, Day T. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*, Chapter 15

- Exercises: #19, #20, #21, #22