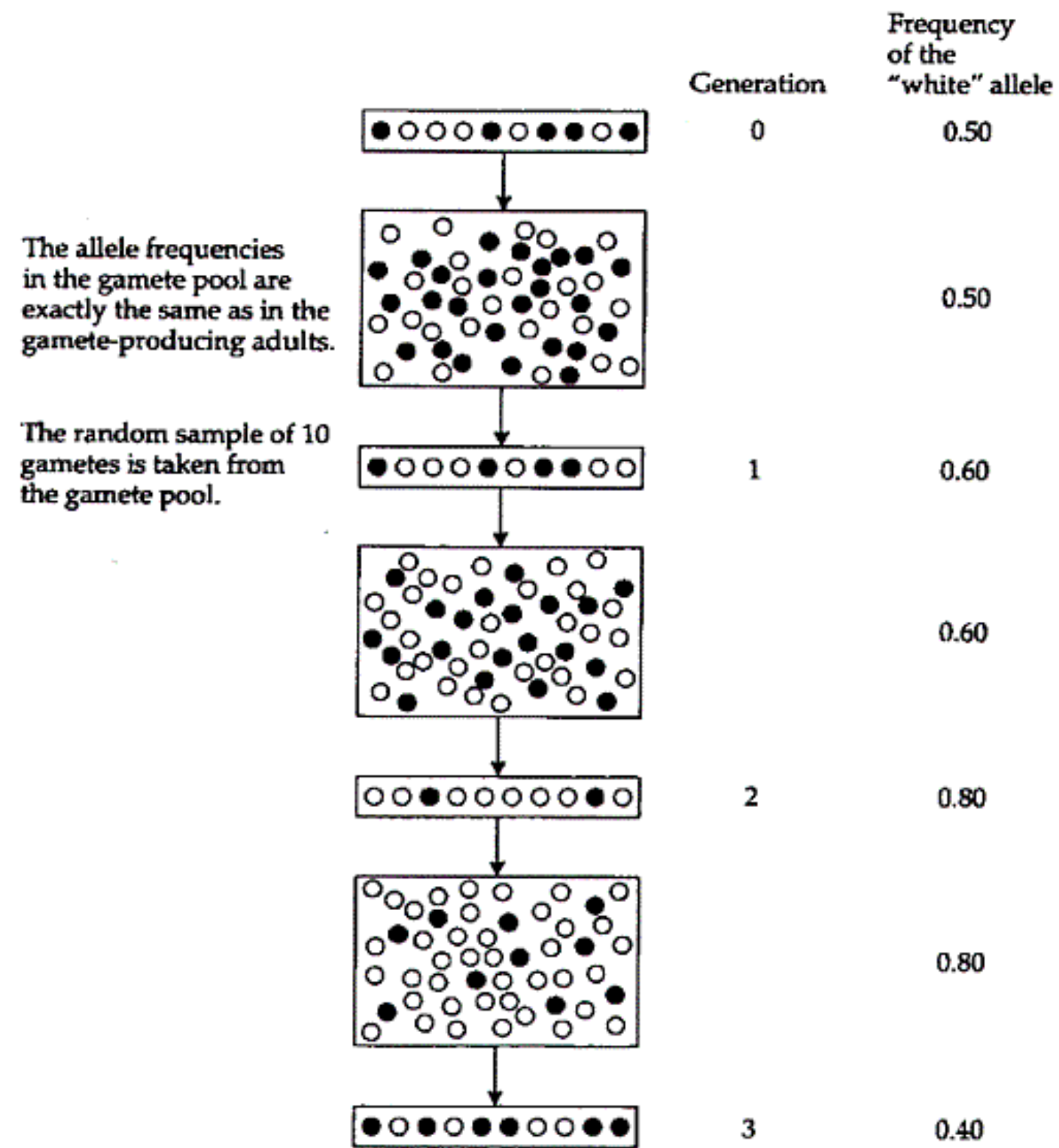


Genetic Drift
and
Neutral Evolution:
A Primer

Luca Peliti
ITP and University "Federico II"

May 22, 2001

Genetic drift



Neutral sampling

Haploid population of constant size N

Two alleles, a and b , for one locus

Assume $W_a = W_b \Rightarrow$ no selection!

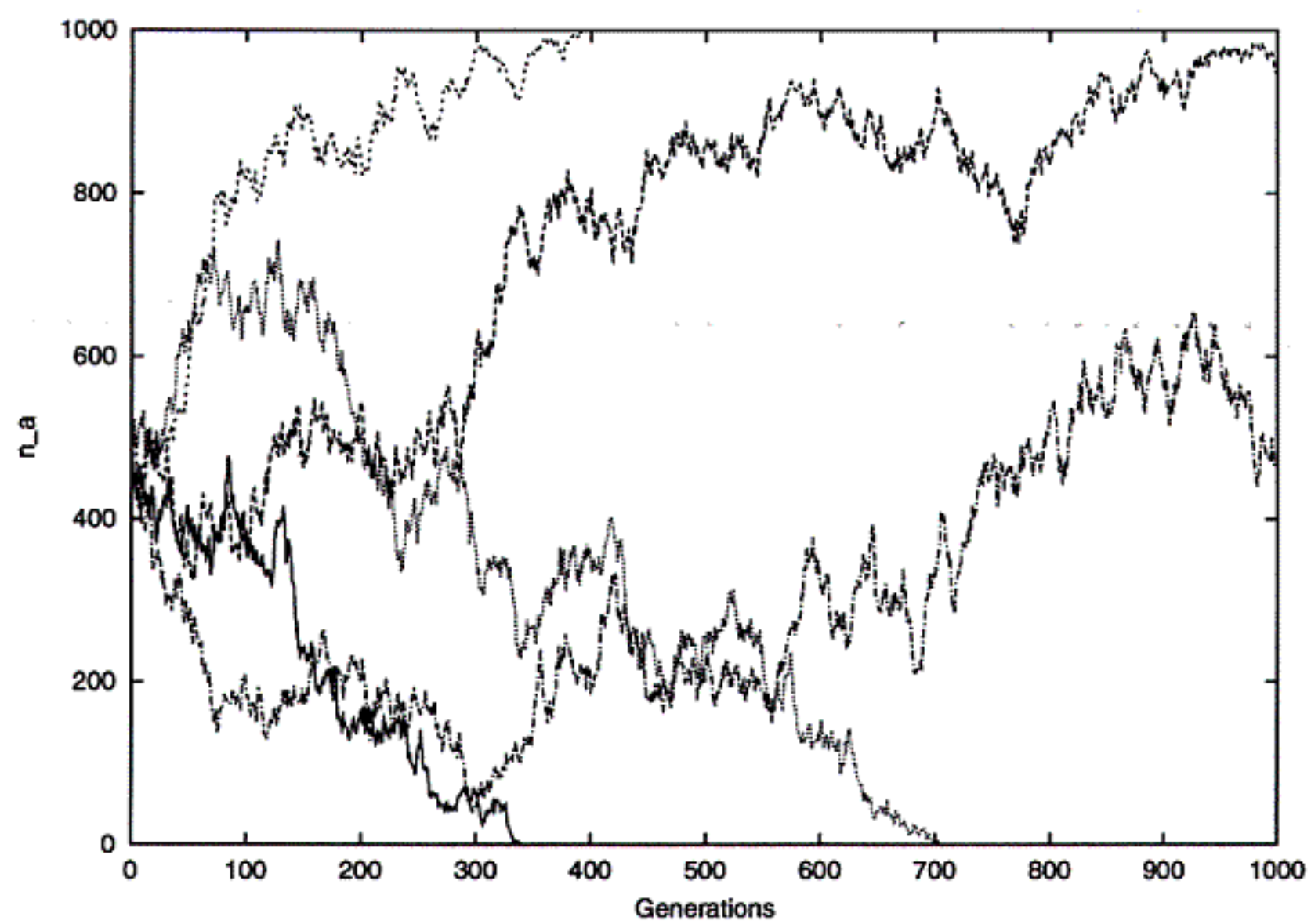
Then

$$P_{n_a}(t+1) = \binom{N}{n_a} (x_a(t))^{n_a} (x_b(t))^{N-n_a}$$

where

$$x_a(t) = \frac{n_a(t)}{N}$$

"Gambler's ruin"



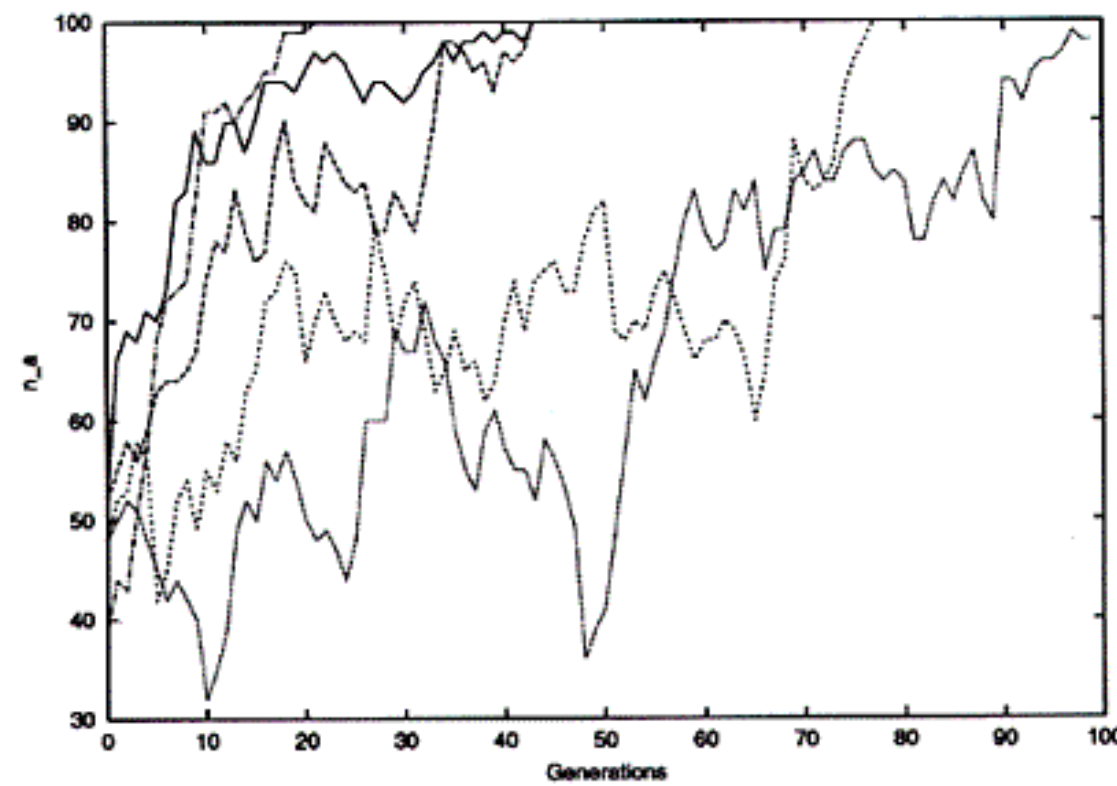
$n_a = N$ \Leftrightarrow **Fixation of a**
 $n_a = 0$ \Leftrightarrow **Loss of a**

Gambler's ruin with selection

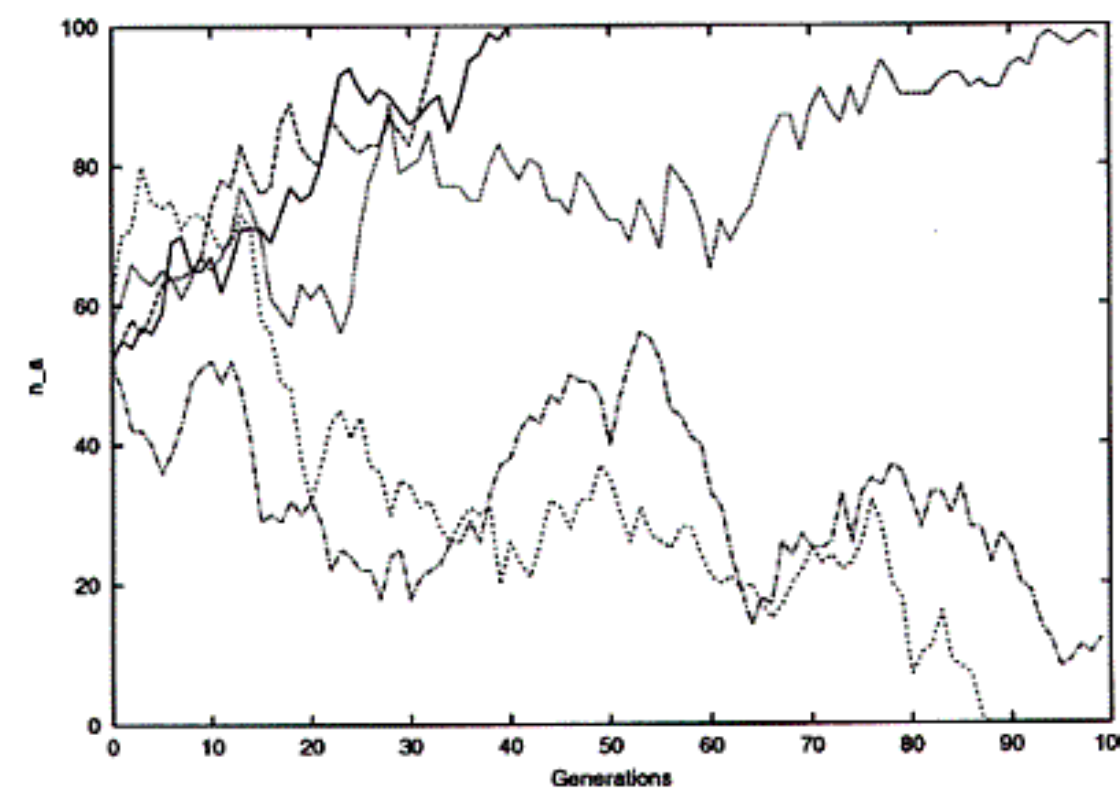
Population of 100 individuals. Relative fitness:

$$W_a = 1, W_b = 1 - s$$

$$s = 0.01$$



$$s = 0.005$$



Kimura's diffusion approximation

Define the gene frequency x at generation t

$$x(t) = \frac{n_a(t)}{N}$$

Define $\mathcal{P}(x, t | x_0, t_0)$ the conditional probability that $x(t) = x$, given $x(t_0) = x_0$ ($t_0 < t$)

Then, for $0 < x < 1$,

$$\frac{\partial}{\partial t} \mathcal{P}(x, t | x_0, t_0) \simeq \frac{\partial^2}{\partial x^2} \{D(x, t)\mathcal{P}\} - \frac{\partial}{\partial x} \{v(x, t)\mathcal{P}\}$$

where

$$D(x, t) \delta t = \frac{1}{2} \left\langle ((x(t + \delta t) - x(t))^2) \right\rangle_{x(t)=x} - \langle x(t + \delta t) - x(t) \rangle_{x(t)=x}^2$$

$$v(x, t) \delta t = \langle x(t + \delta t) - x(t) \rangle$$

Special equations at $x = 0$ and $x = 1$:

$$\frac{d}{dt} \mathcal{P} \Big|_{x=0} = - \left\{ \frac{\partial}{\partial x} (D(x, t)\mathcal{P}) - v(x, t)\mathcal{P} \right\} \Big|_{x=0}$$

$$\frac{d}{dt} \mathcal{P} \Big|_{x=1} = \left\{ \frac{\partial}{\partial x} (D(x, t)\mathcal{P}) - v(x, t)\mathcal{P} \right\} \Big|_{x=1}$$

The backward diffusion equation

We can consider x fixed and x_0 random. Then

$$\frac{\partial}{\partial t} \mathcal{P}(x, t | x_0, t_0) \simeq D(x, t) \frac{\partial^2}{\partial x_0^2} \mathcal{P} + v(x, t) \frac{\partial}{\partial x_0} \mathcal{P}$$

The probability of fixation by time t is given by

$$\phi(x_0, t) = \mathcal{P}(1, t | x_0, 0)$$

Then the *ultimate* fixation probability

$$\phi(x_0) = \lim_{t \rightarrow \infty} \phi(x_0, t)$$

satisfies (assuming a stationary process)

$$D(x_0) \frac{d^2 \phi}{dx_0^2} + v(x_0) \frac{d\phi}{dx_0} = 0$$

with boundary conditions

$$\phi(0) = 0, \quad \phi(1) = 1$$

Random drift in a small finite population

Diploid population of size N

$$P_{n_a}(t+1) = \binom{2N}{n_a} (x_a(t))^{n_a} (x_b(t))^{2N-n_a}$$

↓

$$\langle x(t+1) - x(t) \rangle_{x(t)=x} = 0$$

$$\langle (x(t+1) - x(t))^2 \rangle_{x(t)=x} = \frac{x(1-x)}{2N}$$

Thus

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{1}{4N} \frac{\partial^2}{\partial x^2} \{x(1-x) \mathcal{P}\}$$

with the initial condition

$$\mathcal{P}(x, t=0 | x_0, 0) = \delta(x - x_0)$$

Fixation probability

$$\frac{x(1-x)}{4N} \frac{d^2}{dx_0^2} \phi(x_0) = 0$$
$$\phi(0) = 0 \quad \phi(1) = 1$$

↓

$$\phi(x_0) = x_0$$

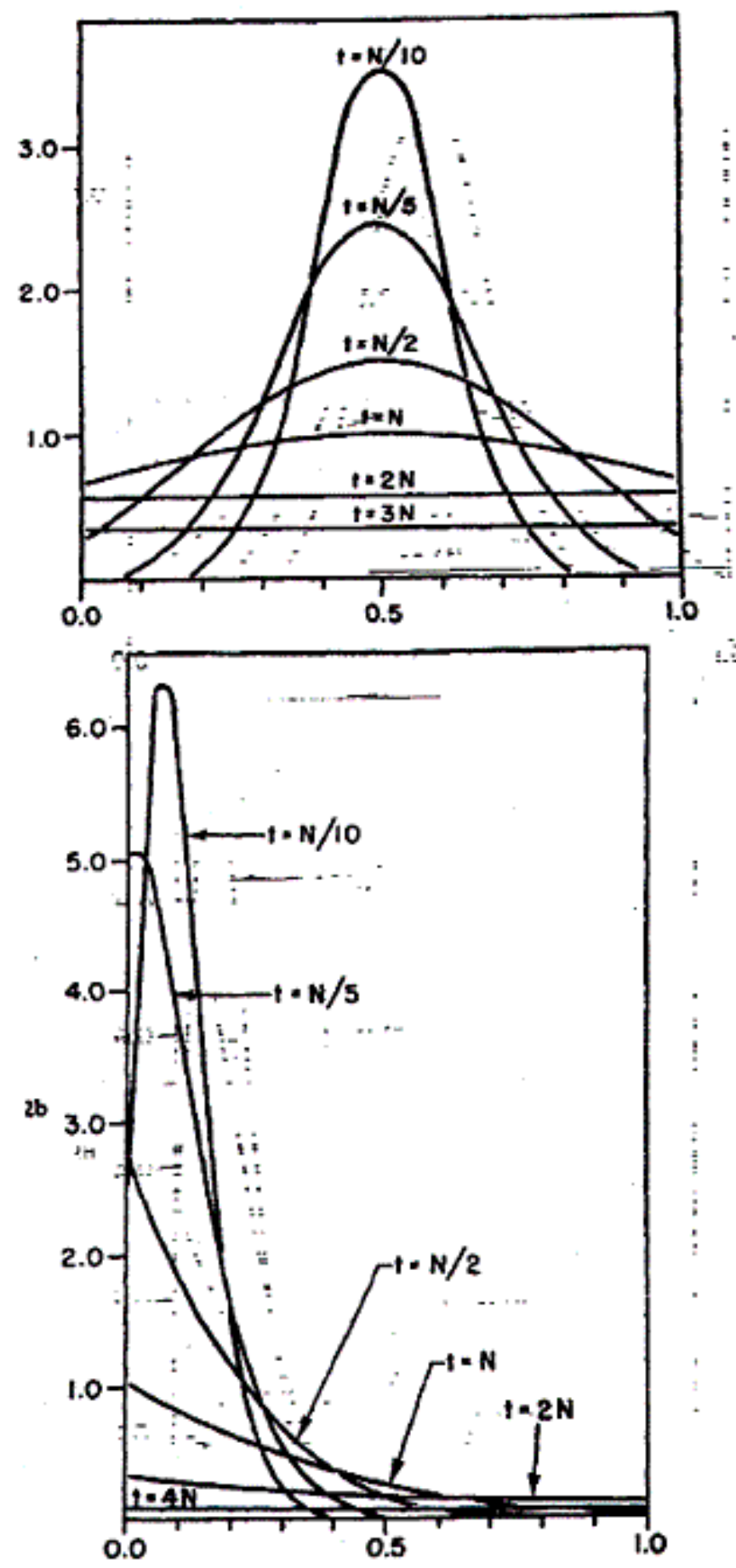
The probability distribution

$$\mathcal{P}(x, t | x_0, 0) = \sum_{\ell=1}^{\infty} C_{\ell} \rho(x_0) \Psi_{\ell}(x) \Psi_{\ell}(x_0) e^{-\lambda_{\ell} t}$$

where

$$\begin{aligned} C_{\ell} &= \frac{(2\ell + 1)}{\ell(\ell + 1)} \\ \rho(x_0) &= 4x_0(1 - x_0) \\ \Psi_{\ell}(x) &= T_{\ell-1}^1(1 - 2x) \quad \text{Gegenbauer polynomial} \\ \lambda_{\ell} &= \frac{\ell(\ell + 1)}{4N} \end{aligned}$$

The change in gene frequencies



Asymptotically for $t \rightarrow \infty$

$$\mathcal{P}(x, t | x_0, 0) \simeq 6x_0(1 - x_0) e^{-t/(2N)}$$

The fraction $H(t)$ of heterozygotes is given by

$$\begin{aligned} H(t) &= \int dx 2x(1 - x) \mathcal{P}(x, t | x_0, 0) \\ &\simeq 2x_0(1 - x_0) e^{-t/(2N)} \end{aligned}$$

The fixation probability by time t can be obtained by solving the backward diffusion equation

One has

$$\begin{aligned} \phi(x_0, t) &= x_0 + \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{2} e^{-\ell(\ell+1)t/(4N)} \\ &\quad \times \left\{ P_{\ell-1}(1 - 2x_0) - P_{\ell+1}(1 - 2x_0) \right\} \end{aligned}$$

where the $P_\ell(x)$ are Legendre polynomials

Multiple alleles

Diploid population of size N , initially n alleles
($\mathbf{a}_1, \dots, \mathbf{a}_n$)

Assume $n \ll N$, and $n_{\mathbf{a}_k}(0) = Nx_k^0$

Then the probability that at time t there are k
alleles with frequency x_k is asymptotically given
by

$$\mathcal{P}_{1,\dots,k}(x_1, x_2, \dots, x_k; t) \simeq (2k-1)! \left(\prod_{\ell=1}^k x_{\ell}^0 \right) e^{-k(k-1)t/(4N)}$$

States with more alleles disappear more rapidly

Coalescence process

$\Pi_k(t)$: Probability that k gene lineages remained independent over t generations

$$\begin{aligned}\Pi_k(t+1) &= \Pi_k(t) \left(1 - \frac{1}{2N}\right) \cdots \left(1 - \frac{k-1}{2N}\right) \\ &\simeq \Pi_k(t) \left(1 - \frac{k(k-1)}{4N}\right) \\ &\simeq \exp\left(-\frac{k(k-1)}{4N}t\right)\end{aligned}$$

Selection and drift

Fitness ratio:

$$\frac{W_a}{W_b} = e^s$$

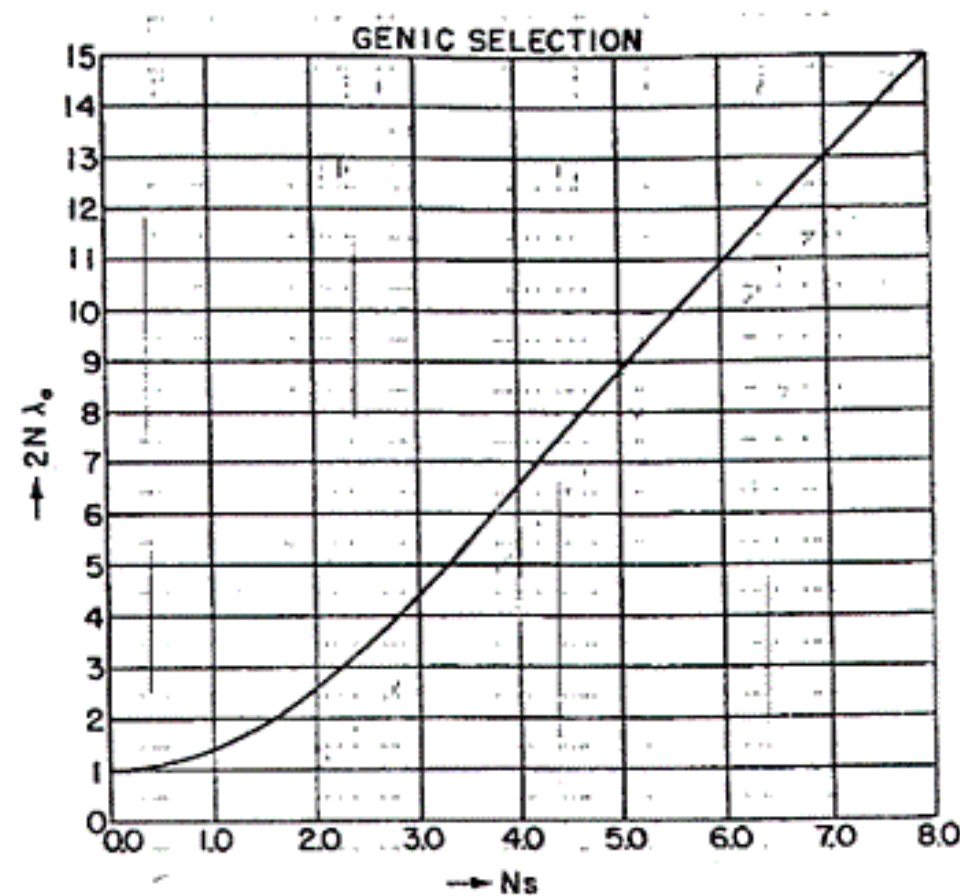
Then

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{1}{4N} \frac{\partial^2}{\partial x^2} \{x(1-x)\mathcal{P}\} - s \frac{\partial}{\partial x} \{x(1-x)\mathcal{P}\}$$

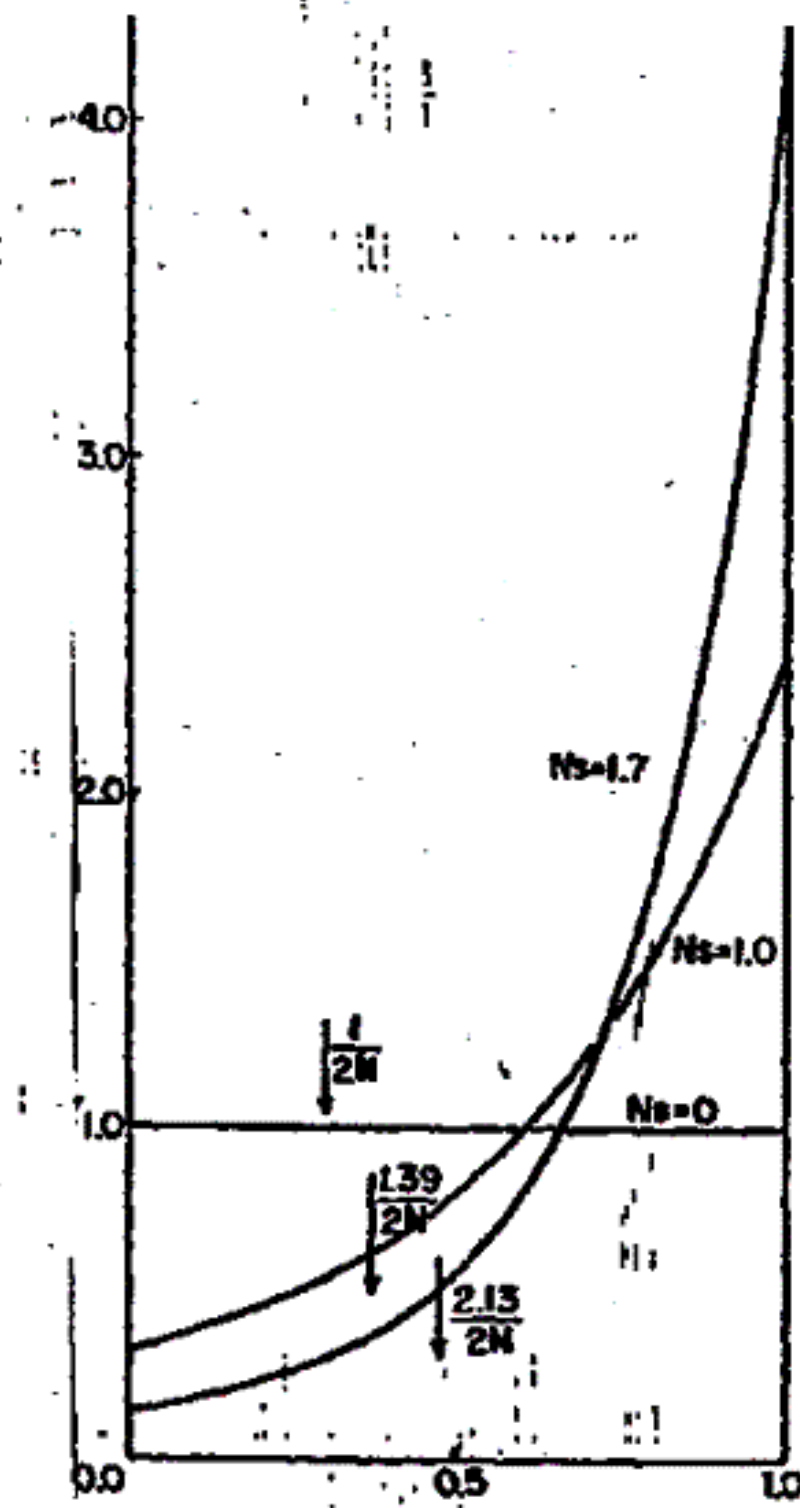
Thus

$$\mathcal{P} = \sum_{\ell=0}^{\infty} C_{\ell} \mathcal{K}_{\ell}(x, x_0) e^{-\lambda_{\ell} t}$$

where $N\lambda_0$ is a function of Ns :



Frequency distribution



Fixation of mutant genes

For $v(x) = sx(1 - x)$ and $D(x) = x(1 - x)/4N_e$ the equation for the fixation probability reads

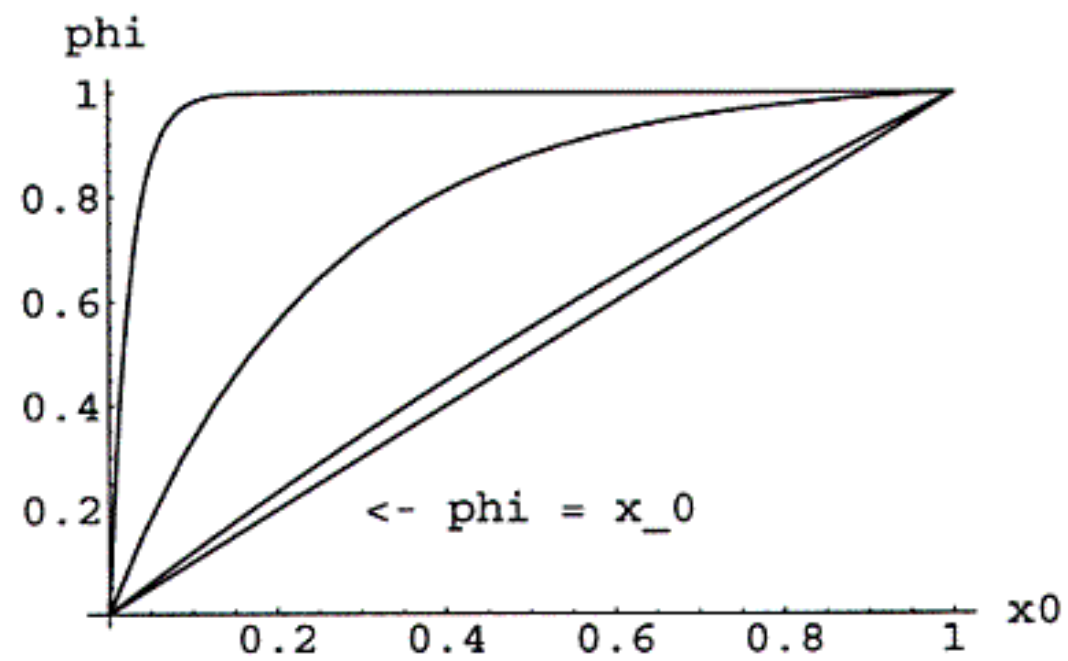
$$\frac{1}{4N_e} \frac{d^2\phi}{dx_0^2} + s \frac{d\phi}{dx_0} = 0$$

N_e : effective population size

Thus

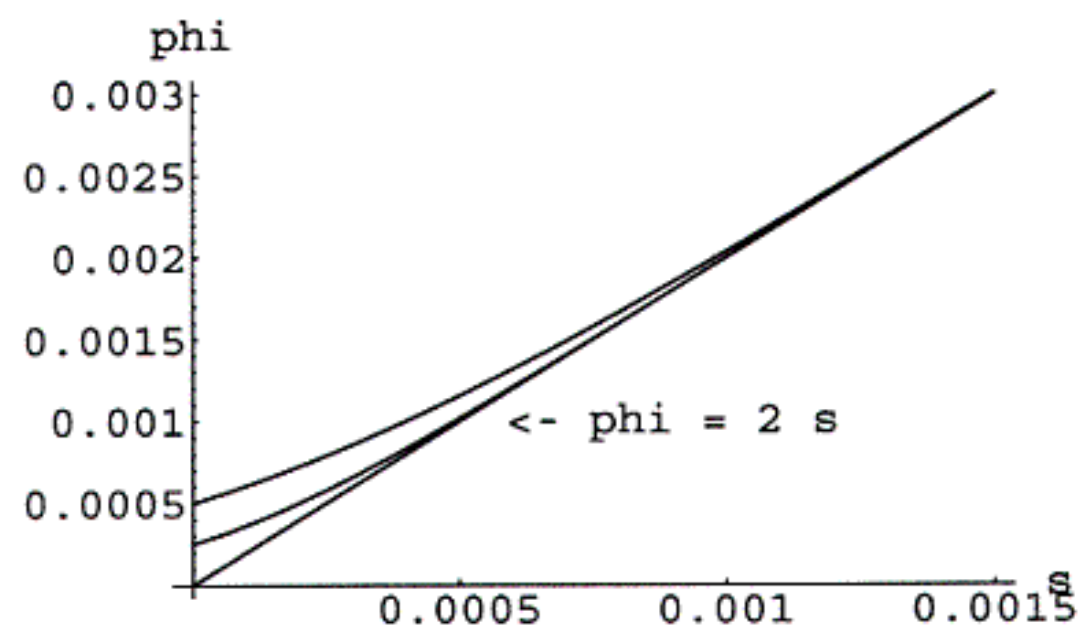
$$\begin{aligned} \phi(x_0) &= \frac{1 - e^{-4N_esx_0}}{1 - e^{-4N_es}} \\ &\approx \begin{cases} x_0 + 2N_esx_0(1 - x_0) & \text{for } 4N_es \ll 1 \\ 4N_esx_0 & \text{for } s \ll 1 \end{cases} \end{aligned}$$

Fixation probability for $s = 10^{-4}$



From top to bottom: $N_e = 10^5, 10^4, 10^3$

Fixation of an advantageous mutant: $x_0 = 1/2N$



Fixation probability vs. s for $N = 1000, 2000$.

Effective population size

Census size: Total number of individuals, N

Effective size: Size of an idealized population that would have the same random sampling effects as the actual population, N_e

If N_m males mate with N_f females, one has

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

If $N_m \ll N_f$ (lek!) then $N_e \simeq 4N_m$

Substitution rate

A mutant gene arises in one copy among the $2N$ available, thus

$$x_0 = \frac{1}{2N}$$

Thus

$$\phi = \begin{cases} 1/(2N), & \text{for neutral mutants} \\ 2N_e s/N, & \text{for advantageous mutants} \end{cases}$$

If the probability of mutation is u per gene and per generation, the fixation rate is

$$\omega = \begin{cases} u, & \text{for neutral mutants} \\ 4uN_e s, & \text{for advantageous mutants} \end{cases}$$

Genetic variability

Consider K alleles: $\mathbf{a}_1, \dots, \mathbf{a}_K$

Mutation rate:

$$\mathbf{a}_i \xrightleftharpoons{u/(K-1)} \mathbf{a}_j \quad \forall i, j$$

Then

$$v_i(\vec{x}) = \langle \delta x_i \rangle_{\vec{x}} = \frac{u}{K-1} (1 - Kx_i)$$

$$D_i(\vec{x}) = \frac{1}{2} \langle \delta x_i^2 \rangle_{\vec{x}} = x_i(1 - x_i)/(4N_e)$$

Assume $K \gg 1$, $2N_e u^2 \ll 1$: Then

$$\mathcal{P}_{\text{eq}}(\vec{x}) \propto \prod_i \left((1 - x_i)^{4N_e u - 1} x_i^{-1} \right)$$

and

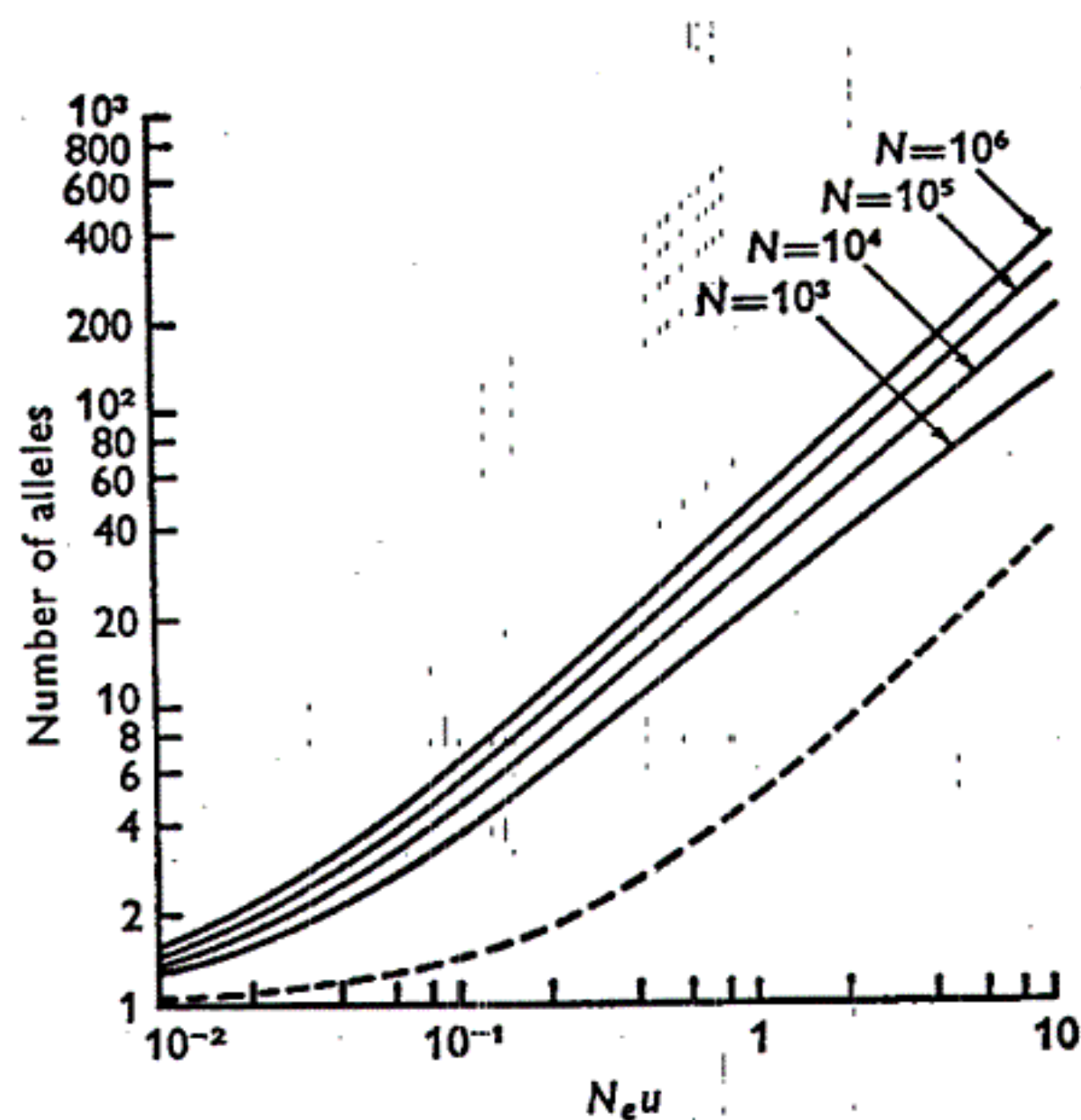
$$\begin{aligned} \langle x_i \rangle_{\text{eq}} &= 1/K \\ \langle x_i^2 \rangle_{\text{eq}} &\simeq \frac{1}{K(4N_e u + 1)} \end{aligned}$$

The *effective* number of alleles is given by

$$n_e = \frac{1}{\sum_i \langle x_i^2 \rangle_{\text{eq}}} \simeq 4N_e u + 1$$

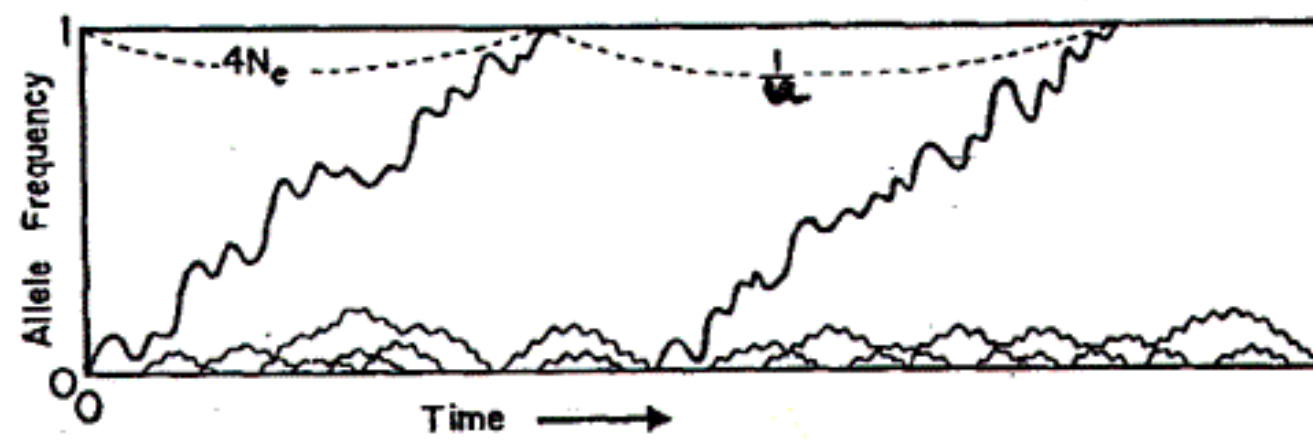
The *average* number of alleles is given by

$$\begin{aligned} n_{\text{av}} &= K (1 - \mathcal{P}^*(0)) \\ &\simeq N_e u \int_{1/(2N)}^1 dx x^{-1} (1 - x)^{4N_e u - 1} \end{aligned}$$

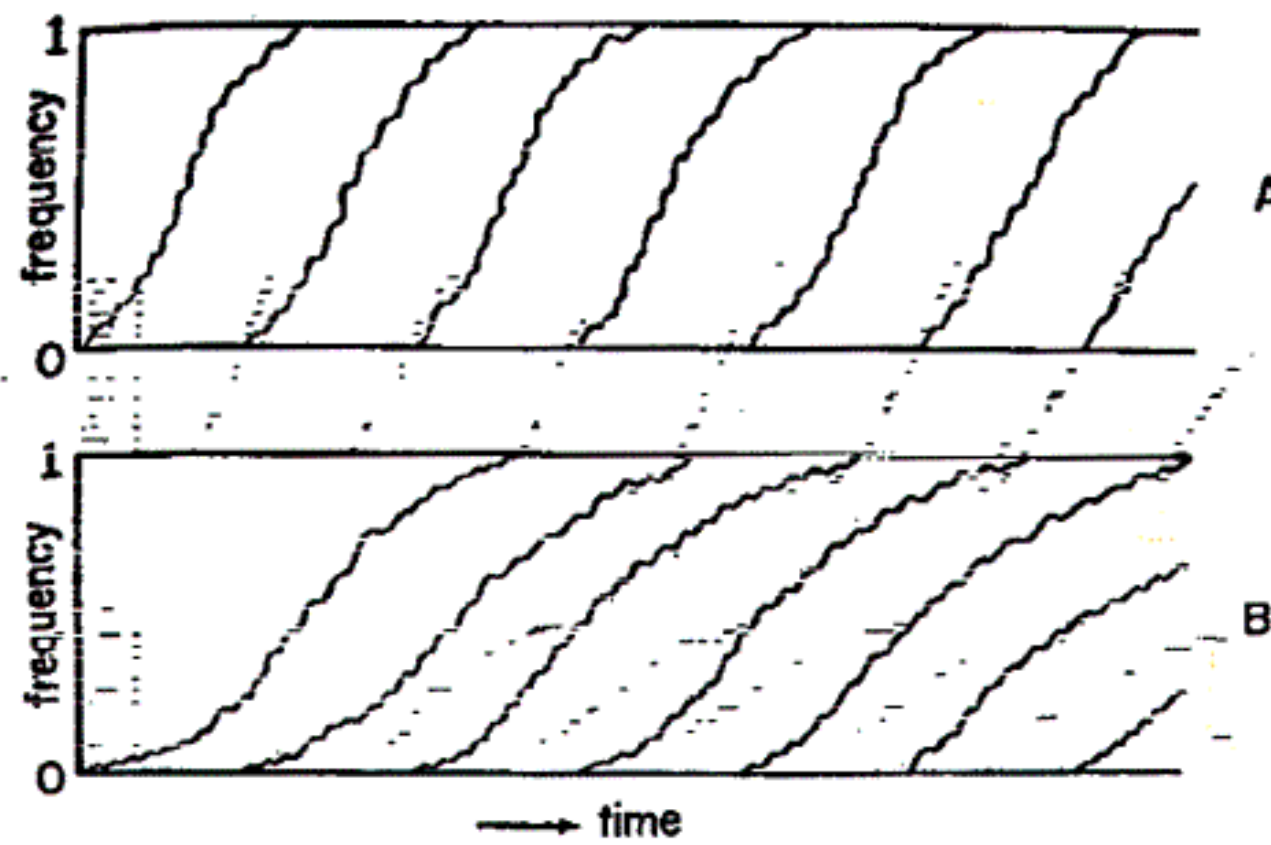


Summarizing

Properties of fixation of neutral mutations

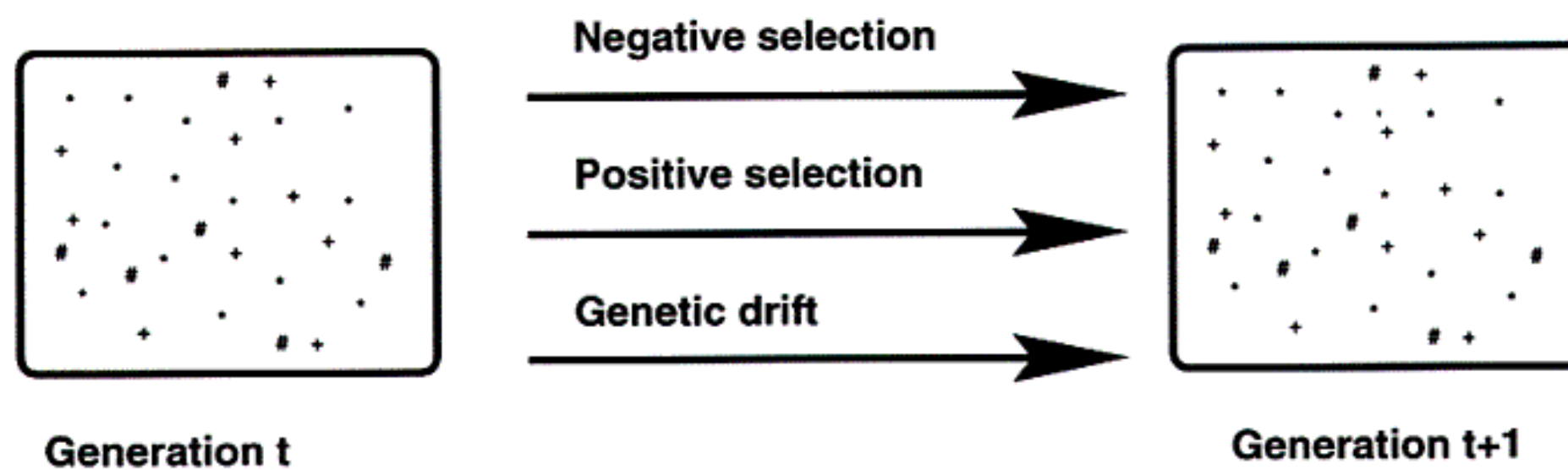


Population-size dependence



Molecular evolution

The three basic mechanisms



What is their relative relevance?

- Substitution rate \lesssim mutation rate: very *large* mutation rate ($\sim 2 \cdot 10^{-8}$ nb/generation in humans (Drake et al. 1995))
- Large mutation rate \Rightarrow large substitutional load, *unless* fitness variation is small
- Most mutations that are substituted are neutral (or only very slightly deleterious)