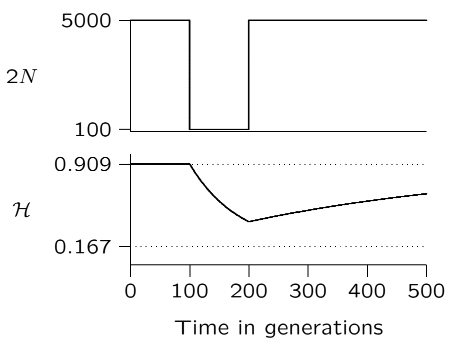
LECTURE: Drift in Populations that Vary in Size

- Heterozygosity is often surprisingly small. Why?
- Urn model assumes N is constant. What if it varies?
- Bottleneck: a temporary reduction in N
- Decline in $\mathcal{H}$ is faster than recovery.
- Effective population size is harmonic mean of N_t

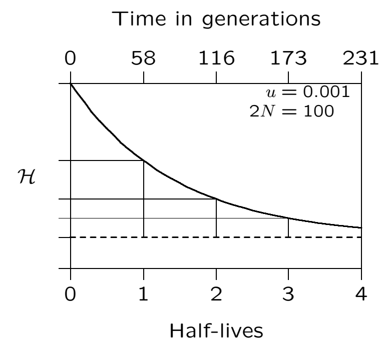
- Harmonic mean is sensitive to small sizes.

A bottleneck in population size



$\mathcal{H}$ declines rapidly, recovers slowly. Why?

What is a half-life?



Why the decline is faster than the recovery

Gene diversity converges toward its equilibrium with a half-life of

$$t_h = \frac{\ln 2}{2u + 1/2N}$$

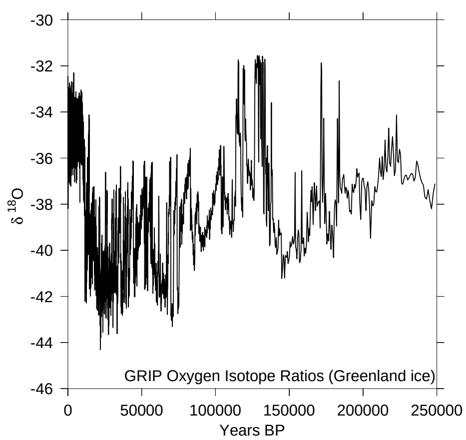
Female population size	Half-life of convergence (generations)	(years)
∞	347	1,041
10^6	346	1,038
10^5	345	1,035
10^4	330	990
10^3	231	693
10^2	58	174
10	7	21

(Assumes $u = 0.001$)

Another problem

We have been assuming that the Elk population was stationary prior to being depleted by human hunting. Is this a reasonable assumption?

Have a look at the record of climate change during the past 250,000 years.



These drastic fluctuations in temperature must have caused fluctuations in population size.

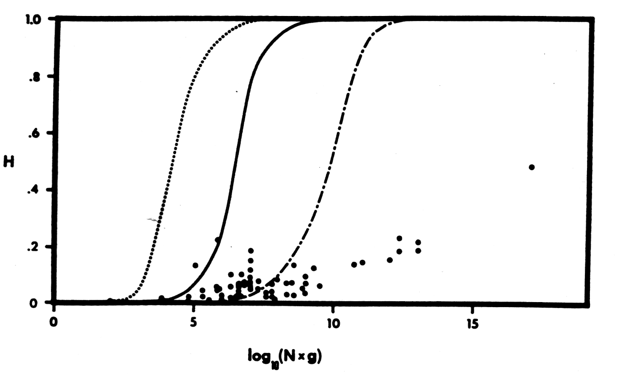


FIG. 1. Relationship between average gene diversity *H* and *Ng*. Dots represent observed values The number of dots is less than 72, because some observed values overlap. Solid line: expectec relationship for neutral alleles with a mutation rate of $v = 10^{-7}$. Dotted line: expected relationshir for overdominant alleles with $v = 10^{-7}$ and $s = 0.001$. Dot-dashed line: expected relationship fo slightly deleterious alleles with $v = 10^{-7}$ and $s = 0.002$. This relationship was obtained by usinç Kimura's (1979) formulas (23) and (24) with $\beta = 0.5$.

For many species, $\mathcal{H}$ is much smaller than would be expected on the basis of their population sizes. Could this be a result of population size bottlenecks during the Pleistocene?

Effective population size, N_e

Goal: Find a value of N that makes our idealized population behave like a more complicated one.

Example: When population size varies, what value of N makes the gene diversity equal to

$$\mathcal{H} = \frac{4N_e u}{4N_e u + 1}?$$

Review: $\mathcal{H}$ in a population of constant size

$$\mathcal{H}_1 = \mathcal{H}_0 \left(1 - \frac{1}{2N}\right)$$

Another generation

$$\mathcal{H}_1 = \mathcal{H}_0\left(1-\frac{1}{2N}\right)$$
$$\mathcal{H}_2 = \mathcal{H}_1\left(1-\frac{1}{2N}\right)$$

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)\\ \mathcal{H}_2 &= \mathcal{H}_1\left(1-\frac{1}{2N}\right)\\ &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)^2\end{aligned}$$

General form

$$\begin{aligned}\mathcal{H}_t &= \mathcal{H}_0\left(1-\frac{1}{2N}\right)^t\\ &\approx \mathcal{H}_0 \exp[-t/2N/]\end{aligned}$$

Note approximation: $1-1/2N \approx e^{-1/2N}$ when N is large.

$\mathcal{H}$ in a population of varying size

$$\mathcal{H}_1 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)$$

where N_0 is population size in generation 0.

Another generation

$$\mathcal{H}_1 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)$$
$$\mathcal{H}_2 = \mathcal{H}_1\left(1-\frac{1}{2N_1}\right)$$

Two generations

$$\mathcal{H}_2 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)\left(1-\frac{1}{2N_1}\right)$$

Two generations again

$$\mathcal{H}_2 = \mathcal{H}_0\left(1-\frac{1}{2N_0}\right)\left(1-\frac{1}{2N_1}\right)$$
$$= \mathcal{H}_0\prod_{i=0}^{1}\left(1-\frac{1}{2N_i}\right)$$

where $\prod$ is the product operator.

General form

$$\mathcal{H}_t = \mathcal{H}_0\prod_{i=0}^{t}\left(1-\frac{1}{2N_i}\right)$$

Approximation

$$\mathcal{H}_t = \mathcal{H}_0\prod_{i=0}^{t}\left(1-\frac{1}{2N_i}\right)$$
$$\approx \mathcal{H}_0\prod_{i=0}^{t-1}\exp[-1/2N_i]$$

How do you simplify a product of exponentials?

Remember rule for product of exponentials

$$e^a e^b = \exp[a+b]$$

Product of exponentials is exponential of sum

$$\mathcal{H}_t \approx \mathcal{H}_0\prod_{i=0}^{t-1}\exp[-1/2N_i]$$
$$= \mathcal{H}_0\exp\left[-\sum_{i=0}^{t-1}\frac{1}{2N_i}\right]$$

Compare results for fixed and varying N

$$\begin{array}{cccc} & \text{Fixed } N & & \text{Varying } N \\ \mathcal{H}_t \approx & \mathcal{H}_0\exp[-t/2N_e] & = & \mathcal{H}_0\exp\left[-\sum_{i=0}^{t-1}\frac{1}{2N_i}\right] \end{array}$$

- N_e is called effective population size.
- It is the constant population size that makes the two sides equal.

The two sides are equal when

$$1/N_e = \frac{1}{t}\sum_{i=0}^{t-1}\frac{1}{N_i}$$

The effective population size, N_e, is the "harmonic mean" of $N_0, N_1, \ldots, N_{t-1}$.

What is N_e good for?

In a population of varying size, heterozygosity at neutral loci is

$$\mathcal{H} = \frac{4N_e u}{4N_e u + 1}?$$

where N_e is the effective population size.

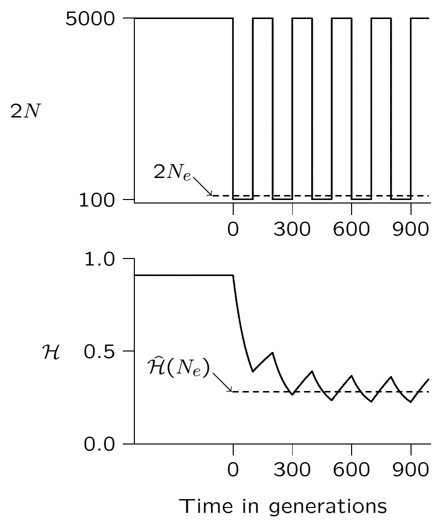
Example

- What is the arithmetic mean of 1, 50, and 100?
- What is the harmonic mean?

Answer

- Arithmetic mean: $(1 + 50 + 100)/3 = 50.3333$.
- Harmonic mean: $1/((1.0 + 1.0/50 + 1.0/100)/3) = 2.9126$.

Harmonic mean is *much* smaller than arithmetic mean.



What do we expect of the Emeryville elk? We would need to look at $\mathcal{H}$ in other populations of large mammals.