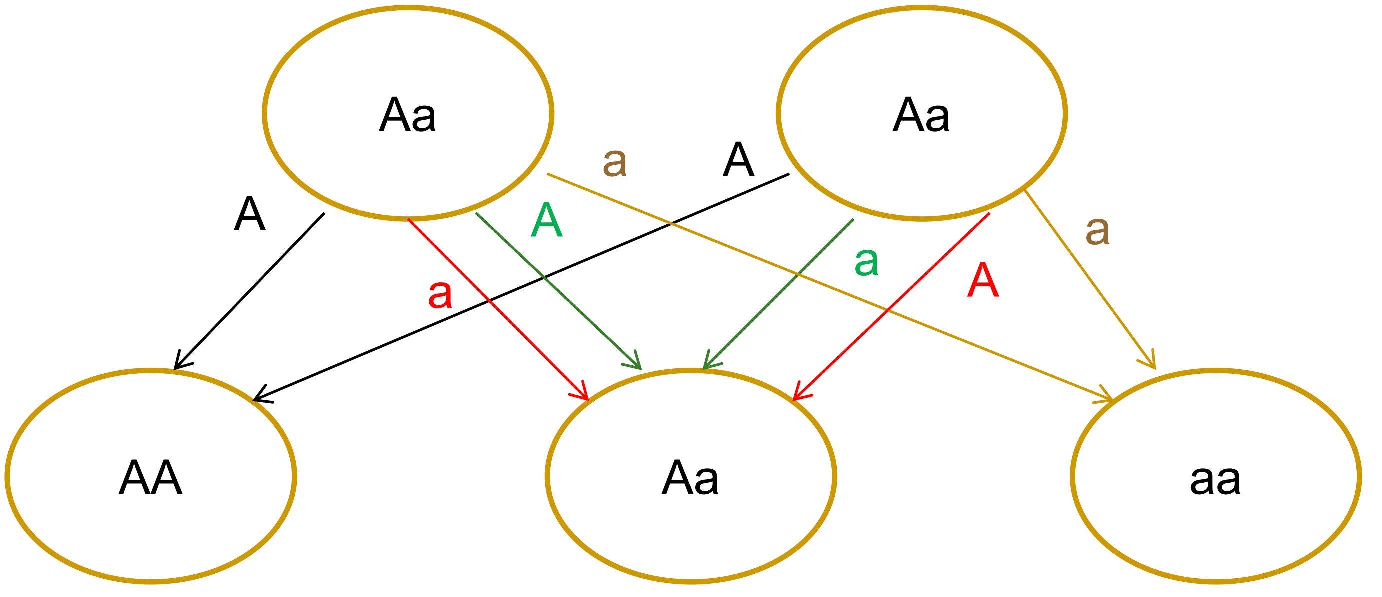
Case Study: Selection of Genes

Background

- Hereditary traits of an organism (e.g. hair and eye colors, colorblindness) are determined by the union of at least one pair of genes.
- The father and mother each contributes one or the other of two alternatives of a gene, frequently denoted by English letters, such as A (dominant gene) and a (recessive gene).
- The resulting zygote may have AA, Aa or aa. An organism with both AA and Aa will exhibit the trait associated with A, while one with aa will exhibit the trait associated with a.

An example



P = 1/4

P = 1/2

P = 1/4

Selection disadvantage

- If a certain fraction s of the zygotes <u>Aa</u> and <u>AA</u> survive to adulthood, while only a fraction $(1-k)s$ of the zygote <u>aa</u> does $(0 < k < 1)$, then the zygote <u>aa</u> is said to have a selective disadvantage, k. That is, the zygote <u>aa</u> is less "fit" than <u>AA</u> or <u>Aa</u> to the environment.
- This disadvantage is often expressed as a percentage, for example, a selective disadvantage of 1% means that $k = 0.01$.

Problem description

- Denote the respective frequencies of A and a genes in the *n*-th generation by p_n and q_n. Given the initial gene frequencies p_0 and q_0, and the selective disadvantage k for aa, how many generations will pass before $q_n < q$, where $q < q_0$ is a positive constant?
- Given p_0, q_0, k, q, solve for n.

Some genetic mathematics

- Assume that the frequencies of genes A and a in an adult population are p and q respectively. If it mates randomly within a generation and no mating between generations, the ratio of genotypes AA, Aa and aa will occur in the offspring population is

$$p^2 : 2pq : q^2$$

- (Hardy-Weinberg ratio)

Some genetic mathematics (cont'd)

- If we know the genotypes AA, Aa, aa occur in the proportion $P:Q:R$, we can compute the gene frequencies p and q (note that $P+Q+R=1$)

$$p = \frac{2PN + QN}{2N} = P + \frac{Q}{2}$$

$$q = 1 - p = R + \frac{Q}{2}$$

Recurrence relation of gene frequency

- Given p_n and q_n for nth generation, the proportion of genotypes AA, Aa and aa in the $(n+1)$st generation will be:

$$p_n^2 : 2p_nq_n : q_n^2$$

- If we take into account the selective disadvantage, the proportion would become

$$p_n^2 : 2p_nq_n : (1-k)q_n^2$$

Recurrence relation of gene frequency (cont'd)

- The proportion can be rewritten as:

$$hp_n^2 : 2hp_nq_n : (1-k)hq_n^2$$

for any positive h.

- Choose h so that

$$hp_n^2 + 2hp_nq_n + (1-k)hq_n^2 = 1$$

- So $P = hp_n^2, Q = 2hp_nq_n$, and $R = (1-k)hq_n^2$
- $p_{n+1} = hp_n^2 + hp_nq_n$
- $q_{n+1} = (1-k)hq_n^2 + hp_nq_n$

Recurrence relation of gene frequency (cont'd)

- Set

$$y_{n+1} = \frac{p_{n+1}}{q_{n+1}} = \frac{p_n^2 + p_n q_n}{(1-k)q_n^2 + p_n q_n}$$

$$= \frac{\left(\frac{p_n}{q_n}\right)^2 + \left(\frac{p_n}{q_n}\right)}{(1-k) + \left(\frac{p_n}{q_n}\right)} = \frac{y_n(y_n+1)}{(1-k)+y_n}$$

- This is the recurrence relation between generations. Note that $y_n = \frac{1-q_n}{q_n}$ or $q_n = \frac{1}{1+y_n}$.

Numerical result

- Assume, for example, $q_0 = 0.99$ and $k = 0.01$. We can use the recurrence relationship to obtain the following results:

n	y_n	q_n
0	0.0101	0.99
1	0.0102	0.9899
2	0.0103	0.9898
…	...	…
1,574	9.0	0.1
…	…	…
10,813	99.0	0.01
…	…	…
101,125	999.0	0.001

The differential approach

- Set $f(n) = y_n$, from previous results we have

$$y_{n+1} - y_n = \frac{y_n(y_n + 1)}{(1 - k) + y_n} - y_n = \frac{k y_n}{y_n + 1 - k}$$

- That is,

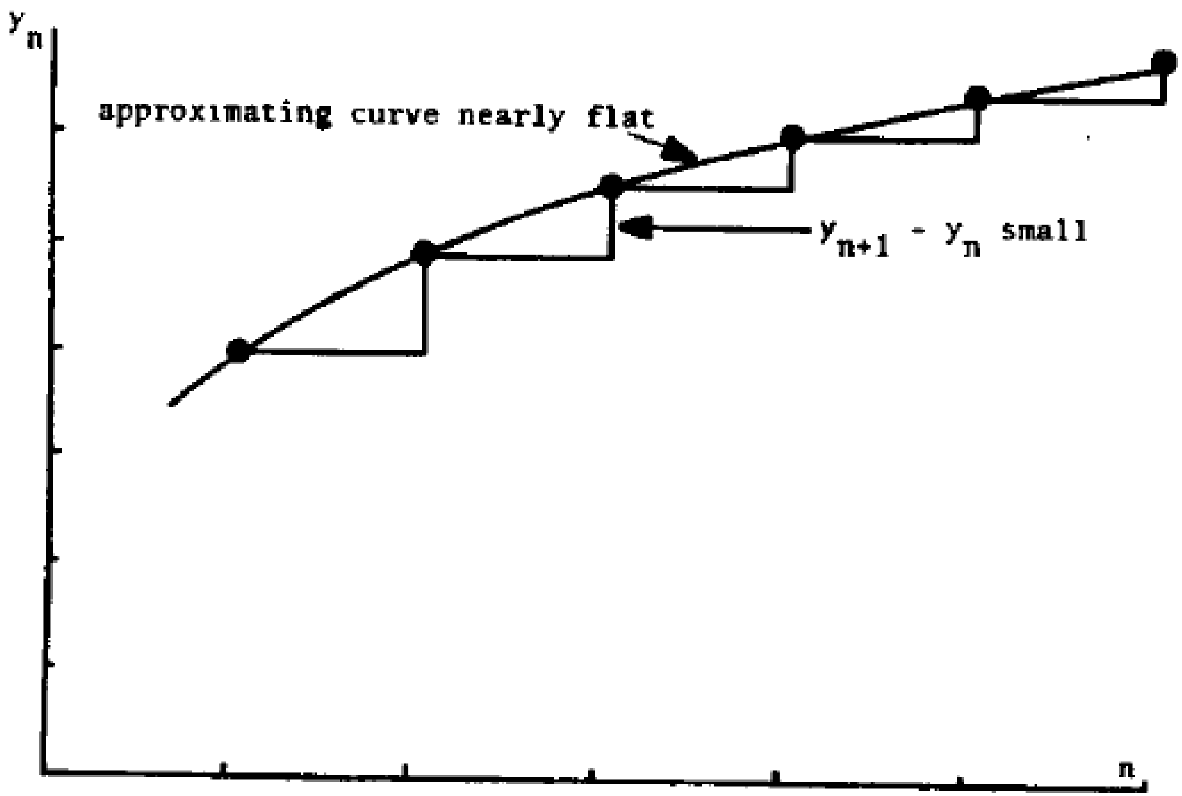
$$f(n+1) - f(n) = k\frac{y_n}{y_n + 1 - k}$$

- Now consider the continuous approximation (substituting n with x), and with a little algebraic operation:

$$\frac{f(x+1) - f(x)}{(x+1) - x} = k\frac{f(x)}{f(x) + 1 - k}$$

The differential approach (cont'd)

- If x is large and/or k is small, then $\frac{f(x+1)-f(x)}{(x+1)-x}$ becomes a good approximation for $f'(x)$.



The differential approach (cont'd)

- Now solve the differential equation

$$f'(x) = k\frac{f(x)}{f(x)+1-k}$$

- This can be transformed to

$$f'(x) + (1-k)\frac{f'(x)}{f(x)} = k$$

- Integrating both sides:

$$f(x) + (1-k)\ln f(x) = kx + C$$

- Use $(0, f(0))$ to solve for C:

$$C = f(0) + (1-k)\ln f(0)$$

The differential approach (cont'd)

- We have

$$f(x) - f(0) + (1-k)\ln\frac{f(0)}{f(x)} = kx$$

- Use back the n, y_n form:

$$y_n - y_0 + (1-k)\ln\frac{y_n}{y_0} = kn$$

- Solve for n:

$$n = \frac{1}{k}\left(y_n - y_0 + (1-k)\ln\frac{y_n}{y_0}\right)$$