Genetic drift

24.10.2007
GE02: day 3 part 2

Yurii Aulchenko
Erasmus MC Rotterdam

Model of a genetic population

- There is a population of n individuals ($2n$ chromosomes)
- A very large number of copies is generated form each chromosomes (gametic pool)
- Next generation is obtained by random sampling of $2n$ chromosomes from this pool

Problem

- Consider a population of 50 people
- One of chromosomes is mutant
- What is the chance that in the next generation the mutation will
  - Disappear?
  - Be still present as single copy?
  - Increase its' frequency?

Solution

- Disappear?
  - $P(k=0) = 0.99^{100} = 0.366$
- Be still present as single copy?
  - $P(k=1) = 100\ 0.01\ 0.99^{99} = 0.37$
- Increase its' frequency?
  - $P(k\geq 2) = 1 - P(0) - P(1\text{ copy}) =$

    $1 - 0.366 - 0.37 = 0.264$

Drift

In a finite genetic populations allelic frequencies are subject to drift (random changes) because of random sampling. The drift is more pronounced with

- Small population size
- Bottleneck effect
  - A large population is reduced very much in size at certain stage
- Founder effects
  - A small group of founders is sampled from large population to start new one

Problem: bottleneck / founder effect

- In a population, mutations of some gene are present with frequencies 0.001 (M1), 0.003 (M2) and 0.005 (M3)
- Due to bottleneck or founder effect, the population is reduced to 50 people (100 chromosomes)
- What is the chance that none of these mutations will be present in founders of the new population?
- What is the chance that all 3 mutations will be presents?

Solution part 1

What is the chance that none of these mutations will be present in founders of the new population?

$$(1 - 0.001 - 0.003 - 0.005)^{100} = 0.405$$

Solution part 2

What is the chance that all 3 mutations will be presents?

Approximate $P(M_1 \geq 1 \ \& \ M_2 \geq 1 \ \& \ M_3 \geq 1)$ by

$$P(M_1 \geq 1)\ P(M_2 \geq 1)\ P(M_3 \geq 1)$$

$P(M_1 \geq 1)\ P(M_2 \geq 1)\ P(M_3 \geq 1) =$

$$= [1 - (1 - 0.001)^{100}]\ [1 - (1 - 0.003)^{100}]\ [1 - (1 - 0.005)^{100}] =$$

$$0.095\ 0.26\ 0.394 = 0.01$$

Drift: very small population

- Consider a “population” made of a single self-pollinating plant
- Initially, the plant is heterozygous (genotype AB)

Problem

What is chance that it will be heterozygous in

- First generation
- 10^{th} generation
- *n*-th generation

After infinite number of generations, what genotypes will be present in the population?

Solution

- What is chance that it will be heterozygous in
  - First generation : (½)
  - 10th generation : $(½)^{10} = 1/1024$
  - *n*-th generation : $(½)^n$
- After infinite number of generations, what genotypes will be present in the population?
- When $n \rightarrow \infty$ then $(½)^n \rightarrow 0$ therefore only AA or BB may be present, with equal chance of ½

Drift

- A population made of $2n$ chromosomes
- k of these are "mutant" (M) and $2n - k$ are "normal" (N). Thus the initial frequency of mutant allele is $p = k/2n$

Drift for 18 chrom. over 19 generations

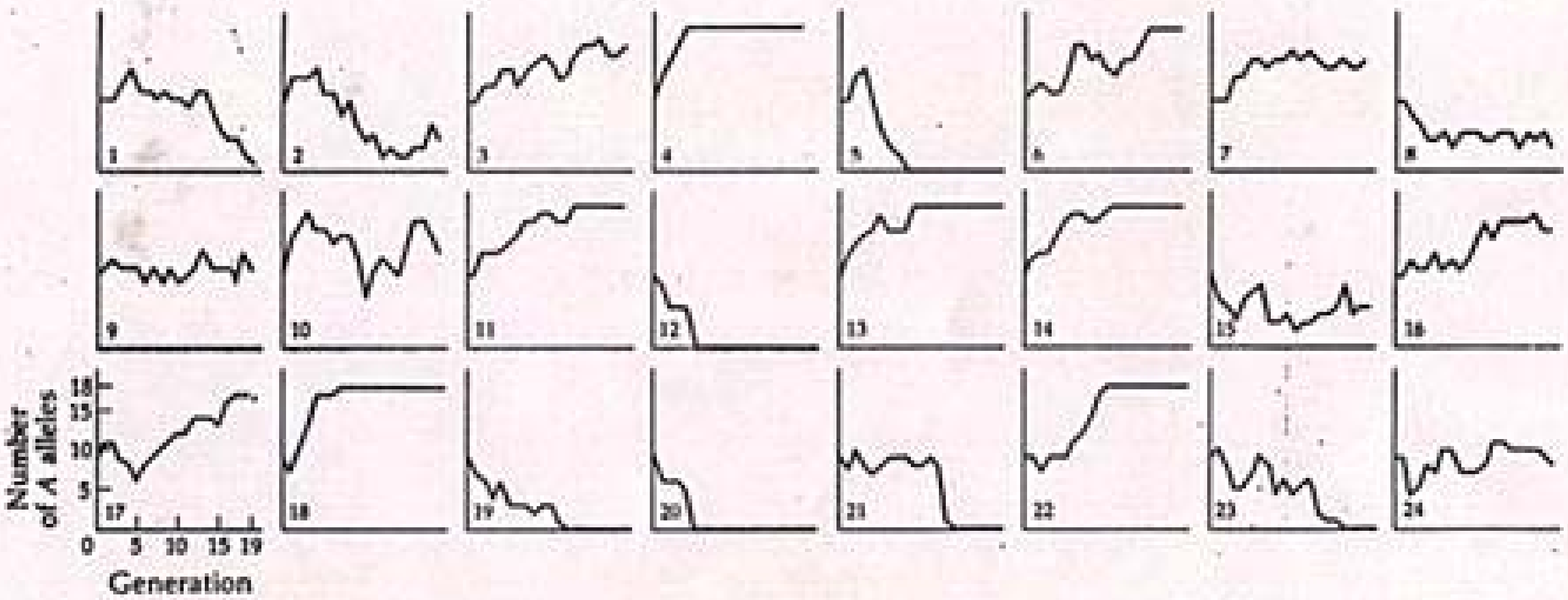

FIGURE 1. Change of allele frequency by random genetic drift over 19 generations in 24 hypothetical populations of size $N = 9$.

Effects of drift

After infinite number of generations, probability that

- Both types are present is 0
- Only M are present is $k/2n = p$
- Only N are present is $(2n - k)/2n = 1 - p$

Time till mutation is lost (fixed)

Expected number of generations before allele is lost is

$$E[t_{lost}] = -\frac{4 \cdot n \cdot p}{(1-p)} \cdot \log_e p; \quad if \;\; p \to 0 \quad then \quad E[t_{lost}] = -2 \cdot \log_e p$$

Expected number of generations before allele is fixed is

$$E[t_{fixed}] = -\frac{4 \cdot n \cdot (1-p)}{p} \cdot \log_e (1-p); \quad if \;\; p \to 0 \quad then \quad E[t_{fixed}] \sim 4 \cdot n$$

After t generations

For an allele with low initial frequency p_0 the loss probability is given by

$$\Pr_{t,p_0}(lost) = \exp\left\{-\frac{4 \cdot n \cdot p_0}{t}\right\}$$

and the variance of frequency is

$$Var_{t,p_0}[p] = p_0 \cdot (1 - p_0) \cdot \left(1 - \left[1 - \frac{1}{2 \cdot n}\right]^t\right)$$

Effective number

- The number discussed is the number of “effective people” describing the population within Generation ⇒ Pool ⇒ Generation model
- It does not directly relate to the number of real people in a population
- *n* is always smaller than the real population size

Estimating effective number

For expanding populations harmonic mean gives a good proxy to effective number

$$\frac{1}{n_e} = \frac{1}{N_{gen}} \cdot \sum_{i=1}^{N_{gen}} \frac{1}{n_i}$$

where N_{gen} is total number of generations and n_i is number of people in i^{th} generation