Lecture 8. Genetic sampling. Wright-Fisher model.

3.1 Genetic sampling

random inheritance of alleles in a finite population

RGD = random change in allele frequencies
due to genetic sampling

Ex 1: D. melanogaster experiment

Fig 7.4, p. 273: $K = 107$ experimental populations
population size $N = 16$
8 males + 8 females chosen at random
One gene with two alleles of equal fitness
brown eye allele $bw = a$
yellow eye allele $bw^{75} = A$

X_t = total number of alleles A in generation t

107 observed scenarios for 20 generations
20 histograms for X_t, $t = 0, 1, \ldots, 19$

Properties of RGD

has no direction, accumulates with time
causes the loss of gen. variability within a population
causes an increase of gen. var. between populations
average allele freq. across populations remains constant

3.2 Wright-Fisher model

WFM is a simple population model on the allele level
assuming random mating and finite population size
Constant generation size
N diploids at each generation
$2N$ haploids (gene copies)
Backward description of the reproduction law
every gene copy picks its parent at random
from $2N$ gene copies in the previous generation
Offspring number distribution $\nu \sim \mathrm{Bin}(2N, \frac{1}{2N})$
Pois(1) approximation: $\mathrm{E}(\nu) = 1$, $\mathrm{Var}(\nu) = 1$

k	0	1	2	3	4	5
$\mathrm{P}(\nu = k)$	0.37	0.37	0.18	0.06	0.02	0.00

Ex 2: RGD simulation

WFM of size 6
random numbers produced by students
six numbers between 1 and 6 per generation
$X_0 = 3$

Allele frequency dynamics

$$\boxed{p_t = \frac{1}{2N}X_t \text{ frequency of allele } A \text{ in generation } t}$$

Fixed initial frequency p_0

independent trajectories of p_t for different populations

Conditional distribution

$X_t \sim \mathrm{Bin}(2N, p_{t-1})$

$\mathrm{E}(p_t|p_{t-1}) = p_{t-1}$, $\mathrm{Var}(p_t|p_{t-1}) = \frac{p_{t-1}q_{t-1}}{2N}$

$\mathrm{Var}(\Delta p) = \frac{pq}{2N}$

Average p_t across populations remains constant

$\mathrm{E}(p_t) = \mathrm{E}(p_{t-1}) = \ldots = \mathrm{E}(p_1) = p_0$

Variation in p_t among populations increases with t

$\sigma_t^2 = \frac{p_0 q_0}{2N} + (1 - \frac{1}{2N})\sigma_{t-1}^2$

Fixation index of RGD

Metapopulation of isolated populations under RGD

fixation index $F_t = \frac{\sigma_t^2}{\bar{p}_t \bar{q}_t} \approx \frac{\sigma_t^2}{p_0 q_0} = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$

$$\boxed{1 - F_t = (1 - \frac{1}{2N})(1 - F_{t-1})}$$

Average heterozigosity across populations (unlinked loci)

$\bar{H}_t = 2\bar{p}_t\bar{q}_t(1 - F_t) \approx 2p_0q_0(1 - F_t)$

In terms of the pedigree inbreeding coefficient

F_t = probab. for two random gene copies at gener. t

to descend from the same gene copy at gener. 0

The rate of RGD in the WFM

$1 - F_t = (1 - \frac{1}{2N})^t$

average heterozygosity $\bar{H}_t \approx 2p_0q_0(1 - \frac{1}{2N})^t$

Literature:

1. D.L.Hartl, A.G.Clarc. Principle of population genetics. Sinauer Associates, 2007.
2. R.Nielson, M. Statkin. An introduction to population genetics: theory and applications, Sinauer Associates. 2013.