# Genetic drift

1  Organisms produce much more gametes than those survive and from offsprings. The process of
2  sampling survival gametes from all ones produced is partly determined by chance, known as genetic drift.

## 1  Wright-Fisher model

Consider a population with the following assumptions: (1) diploid organisms; (2) sexual reporduction; (3) nonoverlapping generations; (4) the gene under consideration has two alleles; (5) identical allele frequencies in males and females; (6) random mating; (7) consistant population size among generations; (8) no immigration; (9) no mutation; (10) no natural selection. These assumptions are identical with those for Hardy-Weinberg model except limited and consistant population size. Let the two alleles $A$ and $a$ have copy number $i$ and $2N - i$ in parental generation, and population size be a constant $N$. The frequency of gametes is $\frac{i}{2N}$ for $A$, and $\frac{2N-i}{2N}$ for $a$. Therefore, the copy number of $A$ in $(n+1)$th generation follows a binomial distribution with parameter $p_n$ and $2N$, *i.e.* the probability that there are $j$ copies of $A$ in the offspring generation is given by

$$T_{ij} = C_{2N}^j (\frac{i}{2N})^j (1 - \frac{i}{2N})^{2n-j} = \frac{(2N)!}{j!(2N-j)!}(\frac{i}{2N})^j (1 - \frac{i}{2N})^{2n-j} \tag{1}$$

4  $T_{ij}$ is the probability that the copy number of $A$ going from $i$ to $j$, known as transition probability. Obvi-
5  ously, the distribution copy number of $A$ is determined by copy number of $A$ in the parental population.
6  Therefore, copy number of $A$ along generations forms a Markov chain. Let the number of generations be
7  large enough and eventually, $A$ will be fixed (copy number reaches $2N$) or lost (copy number drops to 0).
8  The probabilities that $A$ is eventually fixed equals its allele frequency in the founded (0th) generation,
9  since each allele in the founded population has an equal probability to become the ancester of all alleles
10 in the eventual population.

## 2  Identical by descent

Two alleles are identical by descent if they are replicas of a gene present in original generation. Let fixation index $F_t$ be the probability that two alleles randomly sampled from $t$th generation are identical by descent. Therefore, $F_0 = 0$. The allele pool of $(t+1)$th generation is generated by randomly sampling $2N$ alleles from allele pool of $t$th generation with replacement. The event that two alleles randomly drawn from $(t+1)$th allele pool are identical by descent contains two nonoverlapping conditions: (1) the two alleles are replicas of an allele in $t$th generation, which has probability $\frac{1}{2N}$; (2) the two alleles have two distinct parental alleles in $t$th generation, which has probability $1 - \frac{1}{2N}$; and the two parental alleles are identical by descent, which has probability $F_t$. Therefore, the probability that two alleles randomly drawn from $(i+1)$th allele pool are identical by descent, or $F_{t+1}$, is given by

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

Therefore,

$$
\begin{aligned}
1 - F_{t+1} &= 1 - \frac{1}{2N} - (1 - \frac{1}{2N})F_t \\
&= (1 - \frac{1}{2N})(1 - F_t)
\end{aligned}
\tag{3}
$$

Therefore,

$$F_t = 1 - (1 - \frac{1}{2N})^t \tag{4}$$

Therefore, $F_t \to 1, t \to \infty$, *i.e.* when number of generation is large enough, all alleles in the population are identical by descent. Therefore, either $A$ or $a$ achieves an allele frequency of 1, or is fixed. The speed towards fixation is determined by population size $N$: smaller population size, faster fixation.

Consider the probability that two alleles randomly sampled from $t$th generation are not identical by descent, denoted by $H_t$. Therefore, $H_0 = 1$ and $H_t$ is given by

$$
\begin{aligned}
H_t &= 1 - F_t \\
&= (1 - \frac{1}{2N})^t \\
&\approx e^{-\frac{t}{2N}}
\end{aligned}
\tag{5}
$$

As $t$ approaches infinity, both $F_t$ approaches 1 and and $H_t$ approaches to 0 reflects that as time is long enough, all alleles in the population are finally identical by descent, and all individuals are homozygous.

# 3 Effective population size

Effective population size of a real population is the size of an ideal population having the same magnitude of genetic drift as the real population. There are three kinds of magnitude for genetiv drift: (1) change of fixation index $F_t$, corresponding to inbreeding effective size; (2) the change of allele frequency variance, corresponding to variance effective size; (3) the rate of loss of heterozygosity, corresponding to eigenvalue effective size. Inbreeding effective size is the most widely used one and often referred as effective size.

Consider the assumptions of Wright-Fisher model and cancel the assumption of constant population size. Let the population size of $t$th generation be $N_t$ and consider a time period from 0th generation to $(t + 1)$th. Therefore, fixation index $F_t$ is given by

$$
\begin{aligned}
1 - F_{t+1} &= (1 - \frac{1}{2N_t})(1 - F_t) \\
&= (1 - \frac{1}{2N_t})(1 - \frac{1}{2N_{t-1}}) \dots (1 - \frac{1}{2N_0})
\end{aligned}
\tag{6}
$$

To derive effective population size, consider an ideal population with constant size $N_e(t+1)$ and fixation index $F_t$, therefore $F_t$ is given by

$$
1 - F_{t+1} = (1 - \frac{1}{2N_e(t+1)})^{t+1}
\tag{7}
$$

By induction, $N(t+1)$ is given by harmonic mean of real population sizes of $0 - t$th generations, *i.e.*

$$
N(t+1) = \frac{1}{t+1}(\frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N_t})
\tag{8}
$$