Genetic drift

Organisms produce much more gametes than those survive and from offsprings. The process of 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

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

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

2 Diffusion approximation