Population Genetics: The Hardy-Weinberg Model

Sorin Istrail

Department of Computer Science
Brown University, Providence
sorin@cs.brown.edu

September 14, 2010

Outline

Population

In population genetics, the term population does not refer to the entire species, but to a group of organisms of the same species living within a sufficiently restricted geographic area, such that any member can mate with any other member of opposite sex. There are difficulties with this definition. One relates to the fact that geography creates some typically non-random pattern in the spatial distribution of organisms; the members are not uniformly distributed but they are in clusters or colonies, hard to define formally.

Geographic areas that are favorable for habitat are intertwined with areas that are not favorable: towns, cities, rivers, mountains.

Population Substructure

Local interbreeding units of possible large geographical structure are the local units. Such units are of major importance because adaptive evolution takes place through systematic changes in allele frequencies.

Random Mating

- In sexual organisms genotypes are not transmitted to the next generation.

Random Mating

- In sexual organisms genotypes are not transmitted to the next generation.
- Genotype (generation n) $\rightarrow$ Gametes $\rightarrow$ Genotype (generation $n+1$)

Random Mating

- In sexual organisms genotypes are not transmitted to the next generation.
- Genotype (generation n) $\rightarrow$ Gametes $\rightarrow$ Genotype (generation $n+1$)
- The frequency of a genotype in a population is the *genotype frequency*.

Random Mating

- In sexual organisms genotypes are not transmitted to the next generation.
- Genotype (generation n) $\rightarrow$ Gametes $\rightarrow$ Genotype (generation $n+1$)
- The frequency of a genotype in a population is the *genotype frequency*.
- The Genotype of the mating pair determines the genotypes of the offspring.

Random Mating

Definition

The Random Mating Model. Mating pairs have the same frequencies as if they were formed by random collisions between genotypes = the chance that an organism mates with another is based on the genotype frequencies in that population.

Random Mating

- Consider a population with genotype frequencies *AA* 20%, *Aa* 30% and *aa* 50%. Suppose the mating is random: then *AA* females mate with the *AA*, *Aa* and *aa* males in proportion of 20%, 30% and 50% respectively.
- In humans, some mating is random e.g., by blood type; some mating is not random: by height or skin color.

Non-Overlapping Generations

The cycle of birth, maturation, death include the death of all organisms present in each generation before the members of the next generation mature (with is defined as the age of the sexual reproduction).

The Hardy-Weinberg Model for One Locus

- Random mating in the model with non-overlapping generations.
- We assume that the forces that change allele frequencies are negligible, forces such as mutation, migration, natural selection.
- Population must be large in size so that the allele frequencies would not change by changing the sample.

Axioms for the Hardy-Weinberg Model

1. The organism is diploid

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females
6. Mating is random

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females
6. Mating is random
7. Population size is very large (in theory infinite)

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females
6. Mating is random
7. Population size is very large (in theory infinite)
8. Mutation can be ignored

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females
6. Mating is random
7. Population size is very large (in theory infinite)
8. Mutation can be ignored
9. Migration is negligible

Axioms for the Hardy-Weinberg Model

1. The organism is diploid
2. Reproduction is sexual
3. Generations re non-overlapping
4. The gene (locus) under consideration has two allele
5. The allele frequencies are the same for both males and females
6. Mating is random
7. Population size is very large (in theory infinite)
8. Mutation can be ignored
9. Migration is negligible
10. Natural selection does not affect the alleles under consideration

In the Hardy-Weinberg Model the mathematical relation between the allele frequencies and the genotype frequencies is given by:

$AA : p^2$	$Aa : 2pq$	$aa : q^2$

in which

- $p^2, 2pq, q^2$ are the frequencies of the genotypes AA, Aa, aa in zygotes and

In the Hardy-Weinberg Model the mathematical relation between the allele frequencies and the genotype frequencies is given by:

$AA : p^2$	$Aa : 2pq$	$aa : q^2$

in which

- $p^2, 2pq, q^2$ are the frequencies of the genotypes AA, Aa, aa in zygotes and
- p, q are the allele frequencies of A and a in gametes in the previous generation and

In the Hardy-Weinberg Model the mathematical relation between the allele frequencies and the genotype frequencies is given by:

$AA : p^2$	$Aa : 2pq$	$aa : q^2$

in which

- $p^2, 2pq, q^2$ are the frequencies of the genotypes AA, Aa, aa in zygotes and
- p, q are the allele frequencies of A and a in gametes in the previous generation and
- $p + q = 1$

Fundamental Theorem 1

Theorem

The Hardy-Weinberg frequencies are attained after one generation of random mating irrespective of the genotype frequencies in the parental generation.

Fundamental Theorem 2 [Constancy of Allele Frequency in Generations

Theorem

The Hardy-Weinberg law implies the constancy of allele frequencies in every generation and therefore the genotypic composition of the population. This means that in the absence of specific evolutionary forces to change allele frequencies, the mechanism of Mendelian inheritance, by itself keeps the allele frequencies constant and thus preserves genetic variation.

A new notation

- Let us denote the two alleles at fixed locus A_1 and A_2. Suppose that in a generation the frequencies of the three genotypes A_1A_1, A_1A_2, A_2A_2 are denoted by $X, 2Y$ and Z respectively.
- We can calculate again the frequencies of the mating pairs: e.g., for mating type $A_1A_1 \times A_1A_1$ the frequency is X^2; for the mating pair $A_1A_1 \times A_1A_2$ the frequency is $4XY$; and so on.

A new notation

- The Medelian rules indicate that the outcome of the $A_1A_1 \times A_1A_1$ mating is A_1A_1 and similarly we can compute the probabilities of obtaining such a genotype from the other matings types as follows:
-
  - A_1A_1 is obtained from the $A_1A_1 \times A_1A_1$ mating with probability 1
  - A_1A_1 is obtained from the $A_1A_1 \times A_1A_2$ mating with probability $\frac{1}{2}$
  - A_1A_1 is obtained from the $A_1A_2 \times A_1A_2$ mating with probability $\frac{1}{4}$
  - and note that the frequencies of the three mating types are $X^2, 4XY$ and $4Y^2$

A new notation

Therefore, if $X', 2Y', Z'$ are the corresponding three genotypes A_1A_1, A_1A_2, A_2A_2 in the next generation then we have:

- $X' = X^2 + \frac{1}{2}(4XY) + \frac{1}{4}(4Y^2) = X^2 + 2XY + Y^2 = (X+Y)^2$
- $2Y' = \frac{1}{2}(4XY) + \frac{1}{2}(4Y^2) + 2XZ + \frac{1}{2}(4YZ) = 2(X+Y)(Y+Z)$
- $Z' = \frac{1}{4}(4Y^2) + \frac{1}{2}(4YZ) + Z^2 = (Y+Z)^2$

A new notation

$X' = (X+Y)^2$
$2Y' = 2(X+Y)(Y+Z)$
$Z' = (Y+Z)^2$

- In the next generation $X", 2Y", Z"$ we can easily see that: $X" = (X'+Y')^2 = X', Y" = Y', Z" = Z'$ The genotype frequencies obtained in the second generation are maintained in all subsequent generations.

A new notation

Frequencies having this property can be characterized by the relation:

$$\boxed{(Y')^2 = X'Z'}$$

A new notation

- Populations for which $Y^2 = XZ$ are said to have the genotypic frequencies in the Hardy-Weinberg form.
- Let us observe that although there are possible changes in genotypic frequencies between generation 1 and generation 2, the frequency $x = X + Y$ of the allele A_1 does not change between these two generations, nor between any generations following them.

A new notation

- Because $X + 2Y + Z = 1$ only two of the three frequencies $X, 2Y, Z$ are independent (the other follows from them).
- If the $Y^2 = XZ$ holds then only one of the three frequencies is independent (the other two follows from them).
- The most convenient as an independent quantity is the the frequency x of the allele A_1.
- With this notation we now reformulate the Hardy-Weinberg Law.

Hardy-Weinberg Theorem

Theorem

- *Under the assumptions stated, a population, having genotypic frequencies X (of A_1A_1), $2Y$ (of A_1A_2) and Z (of A_2A_2), achieves, after one generation of random mating, stable genotypic frequencies*

$$x^2, 2x(1-x), (1-x)^2$$

  where $x = X + Y, 1 - x = Y + Z$.

- *If the initial frequencies $X, 2Y, Z$ are already of the form $x^2, 2x(1-x), (1-x)^2$ then these frequencies are stable for all generations.*