Hardy-Weinberg Equilibrium

G. H. Hardy (1877 – 1947)
W. Weinberg (1862 - 1937)

Relationship between alleles and genotypes frequencies

Let's consider a population of N diploid individuals at a particular locus with two alleles 0 and 1.

We denote n_{00}, n_{01} and n_{11} the genotypes counts and n_0 and n_1 the allele counts in the population. So, $n_{00} + n_{01} + n_{11} = N$.

We have the following relationships:

=> $n_0 = 2n_{00} + n_{01}$ and $n_1 = 2n_{11} + n_{01}$.

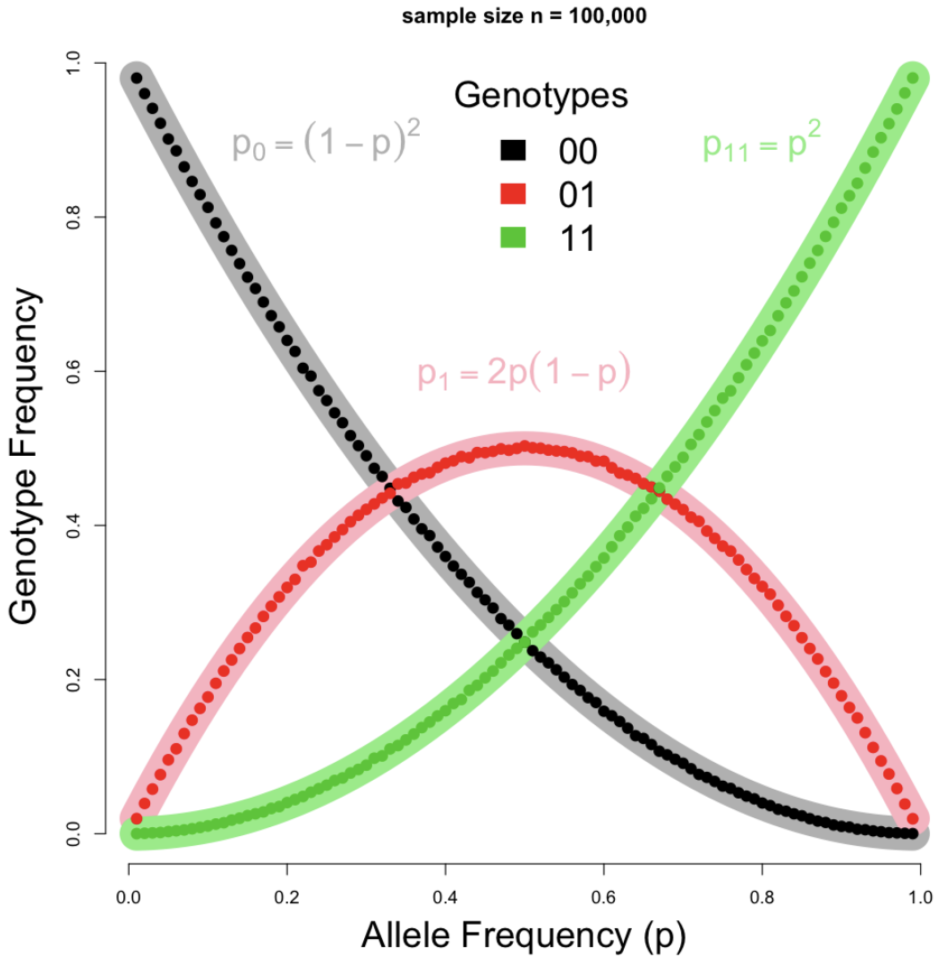
=> $p_0 = n_0 /(2N) = (n_{00} /N) + 0.5(n_{01} /N)$ and $p_1 = 1- p_0$.

$\Rightarrow$ $p_0 = p_{00} + 0.5p_{01}$ and $p_1 = 1- p_0$.

In general, we can not predict the genotype frequencies from the allele frequencies (under-determined).

If genotypes and alleles frequencies are **constant from one generation to the next**, then the population is said to be under Hardy-Weinberg Equilibrium.

If genotypes and alleles frequencies are **constant from one generation to the next**, then the population is said to be under Hardy-Weinberg Equilibrium*.



*This is the diploid / autosomal version of the HWE.

Under which assumption(s) does HWE holds?

Constant allele frequency:
- no migration,
- no mutation,
- no natural selection

Random mating (diploid individuals, sexual reproduction, allele frequencies are the same between sexes)

Large population size

Testing HWE (1/3)

Deviation from HWE can be detected using a χ^2 test with 1 degree of freedom.

Example: Diploid population with the following genotypes counts

Observed			
AA	AB	BB	N_{total}
125	225	150	500

$p_A = (2 \times 125 + 225)/(2 \times 500) = 0.475 \implies p_B = 0.525.$

$p_{AA} = 125/500 = 0.25$, $p_{AB} = 225/500 = 0.45$ and $p_{BB} = 150/500 = 0.3$.

Expectation under HWE: $E[n_{AA}] = p_A^2 \times N_{total} = 112.8125.$

Testing HWE (2/3)

Observed				Expected		
AA	AB	BB	N_{total}	E[AA]	E[AB]	E[BB]
125	225	150	500	112.8	249.4	137.8

Test Statistic

$$\chi^2 = \frac{(125-112.8)^2}{112.8} + \frac{(225-249.4)^2}{249.4} + \frac{(150-137.8)^2}{137.8}$$
$$= 4.78 > 3.84.$$

This example illustrates a **significant** deviation from HWE.

Testing HWE (3/3)

General form of the test statistic

$$\chi^2 = \frac{(n_{AA} - E[n_{AA}])^2}{E[n_{AA}]} + \frac{(n_{AB} - E[n_{AB}])^2}{E[n_{AB}]} + \frac{(n_{BB} - E[n_{BB}])^2}{E[n_{BB}]} \tag{1}$$

$$E[n_{AA}] = N_{total} \times p_A^2$$
$$E[n_{AB}] = N_{total} \times 2p_A p_B$$
$$E[n_{BB}] = N_{total} \times p_B^2.$$

with

$$p_A = (2n_{AA} + n_{AB})/(2N_{total}) \text{ and } p_B = 1 - p_A.$$

Summary

- Population can be characterized by the frequency distribution of alleles and genotypes
- Under certain assumptions, genotype frequencies can be predicted from allele frequencies (Hardy-Weinberg Equilibrium)
- HWE can be extended to sex-linked loci and polyploid individuals
- Deviation from HWE can be used to inform non-random mating (e.g., inbreeding), population history or quality of genetic data (see GWAS lectures)