Original Research Article  

Statistical Model Derivation and Extension of Hardy – Weinberg Equilibrium

Tanveer Ahmed Khan¹*, G. Nanjundan¹, D.M. Basvarajaih² and M. Azharuddin³

¹Department of Statistics, Bangalore University, Bangalore 560056, Karnataka, India  
²Department of Statistics, Dairy Science College, KVAFSU Hebbal, Bangalore, Karnataka, India  
³Department of Genetics, ICAR-NDRI, Adugodi, Bangalore-30, Karnataka, India

*Corresponding author

Abstract

The Hardy-Weinberg equilibrium law states that when a population is in equilibrium, the genotypic frequencies will be in the proportion p²: 2pq: q². In a large random mating hypothetical population where the frequencies of alleles A₁ and A₂ are respectively is p and q, each genotype passes on both alleles with equal frequency over generations in the absence of evolutionary forces (Mutation, migration, and selection). In this paper, the Hardy-Weinberg equilibrium law is derived and extended to the third generation, and the corresponding proportion of frequencies is derived with all mating patterns. The mating frequency matrix is also given. Further, the law is generalized for multiple alleles and generations using binomial expansion.

Keywords  
Hardy-Weinberg law, Genotype frequency, Matings, Binomial expansion, Genetic traits

Introduction

Way back in 1908, a revolutionary contribution has made in genetics and proved statistically by G. H. Hardy and Wilhelm Weinberg, who independently established the principle that the three genotypes A₁A₁, A₁A₂ and A₂A₂ at a bi-allelic locus with allele frequencies p and q = 1 – p are expected to occur in the respective proportions (p²: 2pq: q²) known as Hardy-Weinberg equilibrium (HWE). Some mathematical modeling was formulated based on probability distributions. Fitted models concluded that the gene pool frequencies are inherently stable but that evolutionary forces should be expected in all populations virtually all of the time. Hardy and Weinberg, again they proved the equilibrium stage of a large random mating population. Many geneticists followed them and came to understand that evolution will not occur in a population if the population is large (i.e., there is no genetic drift). All members of the population breed, individuals are mating randomly, and everyone produces the same number of offspring with mutations are negligible, natural selection is not operating in the population, and in the absence of migration in or out of the population. Today, similar studies put forth by many scientists
that the HWE is a prevailing hypothesis used in scientific domains (Ward and Carroll, 2013) ranging from botany (Weising, 2005) to forensic science (Council, 1996) and genetic epidemiology (Sham, 2001; Khoury et al., 2004). The formulation of the theorem will be expressed as follows:

Mendel (1865) rules describe how genetic transmission happens between parents and offspring. Consider a monohybrid cross:

\[ A_1A_2 \times A_1A_2 \]

\[ \frac{1}{4} A_1A_1 \quad \frac{1}{2} A_1A_2 \quad \frac{1}{4} A_2A_2 \]

The Hardy-Weinberg Equilibrium principle

A population with random mating results in an equilibrium distribution of genotypes after only one generation, so that the genetic variation is maintained.

When the assumptions are met, the frequency of a genotype is equal to the product of the allele frequencies.

The Hardy-Weinberg Law (HWL) states that when a population is in equilibrium state, the genotypic frequencies will be in the proportion \( p^2 \), \( 2pq \), and \( q^2 \). In a theoretical population where the frequency of allele \( A_1 \) is \( p \) and the frequency of allele \( A_2 \) is \( q \), each genotype transmit on both alleles that it can posses with equal frequency. Therefore in a population with just two alleles of a gene, the possible combinations as follows:

Random mating | Male | Female |
--- | --- | ---
\( A_1A_1 \times A_1A_2 \) | \( A_1A_1 \) | \( A_1A_2 \) |
\( \frac{1}{4} A_1A_1 \) | \( \frac{1}{2} A_1A_2 \) | \( \frac{1}{4} A_2A_2 \) |

The above mating and proportions show the relationship between the allelic frequencies (\( p \) and \( q \)) and the genotypic frequencies (\( p^2 \), \( 2pq \), and \( q^2 \)), which form the basis of the HWL. For example, the frequency of the genotype \( A_1A_1 \) is \( p^2 \); the frequency of the genotype \( A_1A_2 \) is \( 2pq \).

The HWL states that the allele and genotypic frequencies will remain constant from generation to generation. However, if the population is large, mates randomly, and is free from evolutionary forces (Mutation, migration, and selection). For the above example, it would mean that after taking many generations the frequency of \( A_1A_1 \) is still \( p^2 \) and the frequency of \( A_1A_2 \) is still \( 2pq \).

Stark (2006) demonstrated a model on Clarification of the Hardy–Weinberg Law that HWP can be reached in one round of nonrandom mating with no change in allele frequency.

Stark and Seneta (2012) developed a model which shows that a simple model of non-random mating, which nevertheless embodies a feature of the Hardy-Weinberg Law, can produce Mendelian coefficients of heredity while maintaining the population equilibrium. We can validate this by considering a hypothetical randomly mating population from the table above. To do this, first, consider all the possible matings from every genotypic outcome from table 1.

\[ D^2 + 4DH \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) + 4H^2 \left( \frac{1}{4} \right) \left( \frac{1}{2} \right) = D^2 + DH + \frac{1}{4} D^2 = (D+ \frac{1}{2} H)^2 = p^2 \] (1)

Similarly,

\[ 2DH + 2DR + \left( \frac{1}{2} \right) \left( \frac{1}{4} \right) 4H^2 + \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) 4HR + 2DR + R^2 = 2(D + \frac{1}{2} H)(\frac{1}{2} H + R) = 2pq \]

\[ \left( \frac{1}{2} \right) \left( \frac{1}{4} \right) 4H^2 + \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) 4HR + R^2 = R^2 + HR + \frac{1}{4} H^2 = \left( \frac{1}{2} H + R \right)^2 = q^2 \]
This is in Hardy–Weinberg equilibrium (HWE) after one generation.

**Matrix model**

Let the random mating \(A_1\) with \(A_2\) alleles, then from table 1 we have nine mating combinations, from the parental to offspring generation, as identified by the matrix:

\[
RM = \begin{bmatrix}
A_1 A_1 \times A_1 A_1 & A_1 A_2 \times A_1 A_2 & A_1 A_2 \times A_2 A_2 \\
A_2 A_1 \times A_1 A_1 & A_2 A_2 \times A_1 A_2 & A_2 A_2 \times A_2 A_2 \\
A_2 A_2 \times A_2 A_2 & A_2 A_2 \times A_1 A_2 & A_2 A_2 \times A_2 A_2
\end{bmatrix}
\]

Let the initial mating frequencies

\[
C = \begin{bmatrix}
x_{00} & x_{01} & x_{02} \\
x_{10} & x_{11} & x_{12} \\
x_{20} & x_{21} & x_{22}
\end{bmatrix}
\]

Where \(C\) is the symmetric (Stark and Seneta 2013) i.e., males and females have same frequencies which are denoted by vector \(\{x_0, x_1, x_2\}\).

Let \(C'\) be the transpose of \(C\), that is putting the column vector in row form

\[
C' = \begin{bmatrix}
x_{00} & x_{01} & x_{02} & x_{10} & x_{11} & x_{12} & x_{20} & x_{21} & x_{22}
\end{bmatrix}
\]

Next, we need the Mendel’s coefficients of heredity from table 1 in matrix form are:

\[
M = \begin{bmatrix}
1 & 1/2 & 0 & 1/2 & 1/4 & 0 & 0 & 0 & 0 \\
0 & 1/2 & 1 & 1/2 & 1/2 & 1/2 & 1 & 1/2 & 0 \\
0 & 0 & 0 & 1/4 & 1/2 & 0 & 1 & 1/2 & 1
\end{bmatrix}
\]

The composition of the offspring generation is simply given \(T' = (MC)' \ldots (1)\)

\[
T' = \begin{bmatrix}
x_{00} + \frac{x_{01} + x_{10}}{2} + \frac{x_{11}}{4} & x_{02} & \frac{x_{01} + x_{10}}{2} + \frac{x_{11}}{4}
\end{bmatrix}
\]

If additionally to the conditions of symmetry and sum of all elements equated to unity of \(C\), we also assume that the equilibrium \(H_2^S = 4DR\), that is \(x_{11} = 4x_{02}\) (Stark and Seneta 2013) then the \(T' = (x_0, x_1, x_2)\). If initial population has frequencies \(\{x_0, x_1, x_2\}\), then random mating is expressed as:

\[
C_o = \begin{bmatrix}
(x_0)^2 & x_0 x_1 & x_0 x_2 \\
x_1 x_0 & (x_1)^2 & x_1 x_2 \\
x_2 x_0 & x_2 x_1 & (x_2)^2
\end{bmatrix}
\]

\[
C' = \begin{bmatrix}
(x_0)^2 & x_0 x_1 & x_0 x_2 & x_1 x_0, (x_1)^2 & x_1 x_2 & x_2 x_0, x_2 x_1, (x_2)^2
\end{bmatrix}
\]

Then, applying \(T' = (MC)'\) it will be,

\[
T' = (x_0 + \frac{1}{2} x_1)^2, 2(x_0 + \frac{1}{2} x_1)(x_2 + \frac{1}{2} x_1), (x_2 + \frac{1}{2} x_1)^2
\]

Which in equilibrium \(T' = (p_1^2, 2pq', q_1^2)'\).

**Extension of Hardy Weinberg Equilibrium**

If Random Mating is continued, the second generation is mentioned in table 2. We can first consider all the possible matings from every genotypic outcome above. These matings combinations are listed in Column A.

Next, we are assuming that this population is subject to HWL. Many instances, \(A_1 A_1 \times A_1 A_1\) matings do not occur which often to \(A_2 A_2 \times A_2 A_2\) matings. This would be the frequency of mating between any two genotypes is the product \(T'\). Therefore, the mating frequency of \(A_1 A_1 \times A_1 A_1\), will remain in constant state in equilibrium \(p^3\) or \(p^d\).

Similar results finding were presented in columns C-E are the genotypic frequencies of the next generation. In our example of \(A_1 A_1 \times A_1 A_1\), is equated to 100% of the offspring will
have the genotype $A_1A_1$, so the frequency of that genotype in the next generation is $p^4$.

If the proven HWE is accurate, then the total in Column B should equal the totals of D, C, and E combined, which should come out the same as the frequencies of the original generation. The combined totals of C, D, and E, which makes up the entire population of the next generation, should still result in the same Hardy-Weinberg equation: $p^2+2pq+q^2=1$.

Let the mating frequency matrix from table 2

$$C_o = \begin{bmatrix} (p^2)^2 & 2p^3q & p^2q^2 \\ 2p^3q & 4p^2q^2 & 2pq^3 \\ p^2q^2 & 2pq^3 & q^4 \end{bmatrix}$$

Let $C'$ be the transpose of $C$, that is putting the column vector in row form

$$C' = \{p^4, 2p^3q, p^2q^2, 2p^3q, 4p^3q^3, 2pq^3, p^2q^2, 2pq^3, q^4\}$$

By applying $T' = (MC)'$ obtained from equation (1), we get

$$T' = \{p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6 = (p^2+2pq+q^2)^2 = 1\}$$

Column B: $p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6 = (p^2+2pq+q^2)^2 = 1$

Column C: $p^6 + 4p^5q + 6p^4q^2 + 4p^3q^3 + p^2q^4$

i.e., $p^2 (p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4) = p^2 ((p^2 + 2pq + q^2)^2) = p^2$

Column D: $2p^5q + 8p^4q^2 + 12p^3q^3 + 8pq^4 + 2p^5q$

i.e., $2pq (p^4 + 4p^3q + 6pq^3 + 4pq^3 + q^4) = 2pq ((p^2 + 2pq + q^2)^2) = 2pq$

Column E: $p^4q^2 + 4p^3q^3 + 6p^2q^4 + 4pq^5 + q^6$

i.e., $q^2 (p^4 + 4p^3q + 6pq^3 + 4pq^3 + q^4) = q^2 ((p^2 + 2pq + q^2)^2) = q^2$

We get 3x5 matrix for third generation combination of HWE i.e.

$$C_o = \begin{bmatrix} p^6 & 4p^5q & 6p^4q^2 & 4p^3q^2 & p^2q^4 \\ 2p^5q & 8p^4q^2 & 12p^3q^3 & 8pq^4 & 2pq^5 \\ p^4q^2 & 4p^3q^3 & 6p^2q^4 & 4pq^5 & q^6 \end{bmatrix}$$

From HWE $p^2+2pq+q^2 = 1$, the above proportions are in HWE of the form $p^2+2pq+q^2 = (p+q)^2$, Which in equilibrium $\{p^2, 2pq, q^2\}$. Hence the proof.

**Generalization of Hardy Weinberg Equilibrium (GHWE)**

Ward & Carroll (2013) describes a gene having $r$ alleles $A_1, A_2,..., A_r$ has $r(r+1)/2$ possible genotypes. These genotypes are naturally indexed over a lower-triangular array as $A_1, A_2,..., A_r$. A population is said to be in Hardy-Weinberg Equilibrium (HWE) the law can assumed the following pdf.

If $p_{jk}$ is the relative proportion of genotype $\{A_j, A_k\}$ in the population, and if $\theta_k$ is the proportion of allele $A_k$ in the population, then the system is in HWE if
### Table.1 Model derivation of the Hardy-Weinberg proportions

<table>
<thead>
<tr>
<th>Random Mating</th>
<th>Genotype Frequencies</th>
<th>Mendel’s coefficients of heredity/Conditional probabilities</th>
<th>Mating probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1 \times A_1A_1$</td>
<td>$X_{11}$</td>
<td>$1$</td>
<td>$0$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$X_{11} X_{12}$</td>
<td>$1/2$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$X_{11} X_{22}$</td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_1$</td>
<td>$X_{12} X_{11}$</td>
<td>$1/2$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$X_{12}^2$</td>
<td>$1/4$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$X_{12} X_{22}$</td>
<td>$0$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>$A_2A_2 \times A_1A_1$</td>
<td>$X_{22} X_{11}$</td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$X_{22} X_{12}$</td>
<td>$0$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$X_{22}^2$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

### Table.2 Mating combination for second generation of HWE

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Mating Male x Female</strong></td>
<td><strong>Mating Frequencies</strong></td>
<td><strong>Offspring frequencies</strong></td>
<td><strong>$A_1A_1$</strong></td>
<td><strong>$A_1A_2$</strong></td>
</tr>
<tr>
<td>$A_1A_1 \times A_1A_1$</td>
<td>$p^2 \times p^2 = p^4$</td>
<td>$p^4$</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>$A_1A_1 \times A_1A_2$</td>
<td>$p^2 \times 2pq$</td>
<td>$2p^3q$</td>
<td>$2p^3q$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_1$</td>
<td>$2pq \times p^2 = 4p^3q$</td>
<td>$2p^3q$</td>
<td>$2p^3q$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_1 \times A_2A_2$</td>
<td>$p^2 \times q^2$</td>
<td>--</td>
<td>$2p^2q^2$</td>
<td>--</td>
</tr>
<tr>
<td>$A_2A_2 \times A_1A_1$</td>
<td>$2pq \times 2pq$</td>
<td>$2p^2q^2$</td>
<td>$2p^2q^2$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_2 \times A_2A_2$</td>
<td>$2pq \times 2pq \times 2pq = 4pq^3$</td>
<td>$2pq^3$</td>
<td>$2pq^3$</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2 \times A_1A_2$</td>
<td>$2pq \times q^2$</td>
<td>--</td>
<td>$2pq^3$</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2 \times A_2A_2$</td>
<td>$q^2 \times q^2 = q^4$</td>
<td>--</td>
<td>--</td>
<td>$q^4$</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$(p^2+2pq+q^2)^2 = 1$</td>
<td>$p^2$</td>
<td>$2pq$</td>
<td>$q^2$</td>
</tr>
<tr>
<td></td>
<td>$(p^2+2pq+q^2)^2 = 1$</td>
<td>$2pq$</td>
<td>$2pq$</td>
<td>$q^2$</td>
</tr>
<tr>
<td></td>
<td>$p^2AA + 2pq Aa + q^2 aa = 1$</td>
<td>$p^2$</td>
<td>$2pq$</td>
<td>$q^2$</td>
</tr>
</tbody>
</table>
Table 3: Mating combination for third generation of HWE

<table>
<thead>
<tr>
<th>Type of Mating Male x Female</th>
<th>Mating Frequencies</th>
<th>Offspring frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1 \times A_1A_1 \times A_1A_1$</td>
<td>$p^4 \times p^2$</td>
<td>$p^6 \times p^6$</td>
</tr>
<tr>
<td>$A_1A_1 \times A_1A_1 \times A_1A_2$</td>
<td>$p^4 \times 2pq$</td>
<td>$6p^5 \times q$</td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_1 \times A_1A_1$</td>
<td>$2pq \times p^2 \times p^2$</td>
<td>$15p^4 \times q^2$</td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_2 \times A_1A_1$</td>
<td>$p^2 \times p^2 \times q^2$</td>
<td>$20p^3 \times q^3$</td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_2 \times A_1A_2$</td>
<td>$2pq \times 2pq \times 2pq$</td>
<td>$15p^2 \times q^4$</td>
</tr>
<tr>
<td>$A_2A_2 \times A_2A_2 \times A_2A_2$</td>
<td>$2pq \times q^2 \times q^2$</td>
<td>$6pq^5$</td>
</tr>
<tr>
<td>$A_2A_2 \times A_2A_2 \times A_2A_2$</td>
<td>$q^2 \times q^2 \times q^2$</td>
<td>$q^6$</td>
</tr>
</tbody>
</table>

$p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6 = (p+q)^6 = 1$

Multiple alleles

The expected genotypic array under Hardy-Weinberg equilibrium for two alleles say $A_1A_1$ and $A_2A_2$ is $p^2$, $2pq$, and $q^2$, which form the terms of the binomial expansion $(p+q)^2$. To generalize to more than two alleles, one...
need only add terms to the binomial expansion and thus create a multinomial expansion. For example, with alleles $A_1$, $A_2$, and $A_3$ with frequencies $p$, $q$, and $r$, the genotypic distribution should be $(p+q+r)^2$, or homozygote will occur with frequencies $p^2$, $q^2$, and $r^2$, and heterozygote will occur with frequencies $2pq$, $2pr$, and $2qr$. Further, if we have multiple alleles $A_1$, $A_2$, ..., $A_k$ with genotype probability frequencies $x_1$, $x_2$, ..., $x_k$ such that $\sum x_k = 1$, then the multinomial expansion is given as

$$\sum_{x_1,x_2,\ldots,x_k}^{2} \binom{k}{x_1,x_2,\ldots,x_k} x_1^2 x_2^2 \ldots x_k^2$$

Multiple generation / loci

If males and females each have the same two alleles $P$ and $Q$, then genotypes will be distributed as a binomial expansion in the frequencies $p^2$, $2pq$, and $q^2$. From the above derivations the Hardy–Weinberg equilibrium can be extended to include, among other cases, multiple alleles and multiple generations. i.e., for the first generation with probabilities $P$ and $Q$ it is $(P+Q)^2 = p^2 + 2pq + q^2 = 1$ with four genotypes. At second generation it is $((P+Q)^2)^2 = (P+Q)^4 = p^4 + 4p^3Q + 6p^2Q^2 + 4pq^3 + q^4$ with mating combination of $3^2 = 9$ genotypes. Third generation it is $((P+Q)^4)^2 = (P+Q)^6 = p^6 + 6p^5Q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6 = 1$ with mating combination $3^3 = 27$ genotypes and therefore for the $n^{th}$ generation we generalize using binomial distribution with $3^n$ Combination genotypes and the distribution pattern of F2 genotypes is $((P+Q)^2)^n$:

$$(P + Q)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} q^k p^{2n-k}$$

$$= \binom{2n}{0} p^{2n} + \binom{2n}{1} p^{2n-1} Q + \ldots + \binom{2n}{k} p^{2n-k} Q^{k} + \ldots + \binom{2n}{2n} p^{2n} Q^{2n}$$

With matrix of size 3 x (2n-1) rank.

Edwards (2008) accounted G. H. Hardy’s role in establishing in the existence of “Hardy–Weinberg equilibrium.” Stark A E (2006) demonstrated a model on Clarification of the Hardy–Weinberg Law that HWP can be reached in one round of nonrandom mating with no change in allele frequency. Crow (1988) made remarks that ever since its discovery in the early 1900s, the Hardy–Weinberg law has been a subject of intense consideration and a powerful research tool in population genetics. Stark (2006) reviewed the most basic law of population genetics, which is attributed to Hardy (1908) and Weinberg (1908), which is poorly understood by many scientists who use it routinely. As per the present study, HWE derived and extended from the second generation to the third generation with all possible mating including matrix form. Further, the law is generalized for multiple alleles and multiple generations using binomial expansion. At second generation it is $((P+Q)^2) = (P+Q)^4 = p^4 + 4p^3Q + 6p^2Q^2 + 4pq^3 + q^4$ with mating combination of $3^2 = 9$ genotypes for third generation it is $((P+Q)^4)^2 = (P+Q)^6 = p^6 + 6p^5Q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6 = 1$ with mating combination $3^3 = 27$ genotypes and therefore for the $n^{th}$ generation we generalize using binomial distribution with $3^n$ Combination genotypes and the distribution pattern of F2 genotypes is $((P+Q)^2)^n$ With matrix of size 3 x (2n-1) rank.

References


How to cite this article: