A Simple Proof of the Hardy-Weinberg Law

YAP Von Bing (stayapvb@nus.edu.sg)
Department of Statistics and Applied Probability
National University of Singapore
May 11, 2014

Keywords: Hardy-Weinberg, Mendel, conditional independence, Yule’s paradox.

1 Introduction

Suppose that in a population, an autosomal locus has $k$ alleles $a_1, \ldots, a_k$, with proportions $\text{pr}(a_i) = p_i$. Since the $k(k+1)/2$ genotypes outnumber the alleles, the allele distribution does not determine the genotype distribution in general. An important exception is the Hardy-Weinberg equilibrium, where the genotype proportions are

\[ \text{pr}(a_i a_j) = \begin{cases} p_i^2, & i = j \text{ (homozygote)} \\ 2p_i p_j, & i < j \text{ (heterozygote)} \end{cases} \]  

In 1908, Hardy and Weinberg independently showed that in the case $k = 2$, (1) holds in a large offspring population generated by random mating, assuming that mothers and fathers have the same genotype distribution, and there is no mutation, selection or migration. In fact, the same conclusion follows if mothers and fathers have the same allele distribution; their genotype distributions can be different. “Equilibrium” refers to the fact that under random mating among the children, the grandchildren genotype distribution is still (1), etc.

Hardy’s proof assumes $k = 2$, as is commonly presented in textbooks ([Ewe] pages 3–6, [Ham] pages 17–19). A proof for arbitrary $k$ is in [Edw] (pages 6–7). All these arguments derive the probability of a genotype by summing over and conditioning on relevant mating types. Here I present a new proof which bypasses the algebra.

2 The new proof

Let $M$ and $F$ be the random maternal and paternal genes in a child. Then we have

**Lemma.** (Gametic independence) Under random mating, $M$ and $F$ are randomly chosen from the respective pools of genes in the mothers and fathers, and they are independent.

**Proof.** Suppose there are $n$ mothers. Put their genes as rows in an $n \times 2$ matrix. Random mating means to to choose a row at random, and by Mendel’s First Law, a gene is chosen at random from this row. Hence, the maternal gene is chosen at random from the $2n$ genes. The process is analogous for the fathers, and the two processes are independent. $\square$

Now (1) follows easily. Since \( \text{Pr}(a_i) = p_i, 1 \leq i \leq k \), \( \text{Pr}(M = a_i) = \text{Pr}(F = a_i) = p_i \). By independence, \( \text{Pr}(M = a_i, F = a_j) = p_i p_j, 1 \leq i, j \leq k \). For the homozygotes, \( i = j \), so \( \text{Pr}(a_i a_i) = \text{Pr}(M = a_i, F = a_i) = p_i^2 \). For the heterozygotes, \( i < j \), and \( \text{Pr}(a_i a_j) = \text{Pr}(M = a_i, F = a_j) + \text{Pr}(M = a_j, F = a_i) = 2p_i p_j \).

The Lemma does not assume that mothers and fathers have a common allele distribution. Suppose that the allele distribution in the mothers is \( p_{m1}, \ldots, p_{mk} \), and that in the fathers is \( p_{f1}, \ldots, p_{fk} \). Then equilibrium takes two generations. The Lemma implies that the offspring genotype distribution is

\[
\text{Pr}(a_i a_j) = \begin{cases} 
p_i m p_i f, & i = j \text{ (homozygote)} \\
 p_i m p_j f + p_i f p_j m, & i < j \text{ (heterozygote)} 
\end{cases}
\]

Since this holds in both female and male offspring, their offspring genotype distribution is in Hardy-Weinberg equilibrium, with \( \text{pr}(a_i) = (p_i m + p_i f)/2 \).

3 Discussion

Why is the new proof simpler? A progeny comes about by three steps:

1. Sampling of parents.
2. Sampling of gametes, given parents.
3. Fusion of gametes.

Mendel’s First Law combines steps 2 and 3 to obtain the genotype distribution from each mating type. Bringing in step 1 necessitates summing over mating types, hence the algebraic complexity. Since 1 and 2 occur independently within the maternal and paternal populations, combining them first easily demonstrates gametic independence.

The Yule’s paradox [Yul] (or Simpson’s paradox) refers to the phenomenon that relationships between variables in subgroups can be reversed when the subgroups are combined. In particular, two variables can be conditionally independent in all subgroups, but unconditionally dependent. From this perspective, the Hardy-Weinberg Law is a “counter-example”: \( M \) and \( F \) are conditionally independent given any mating type, but they are unconditionally independent. This is a clue to the existence of a simple proof. In the old proof, gametic independence is inferred from (1); for a proof, see [Edw].

Acknowledgement. I thank Terry Speed for valuable comments.

References

[Edw] Edwards, AWF. *Foundations of Mathematical Genetics 2e*, Cambridge University Press (2000).
[Ewe] Ewens, WJ. *Mathematical Population Genetics 2e*, Springer (2004).
[Ham] Hamilton, MB. *Population Genetics*, Wiley-Blackwell (2009).
[Har] Hardy, GH. Mendelian proportions in a mixed population, *Science* 28:49–50 (1908).
[HCC] Hartl, DL and Clark, AG. *Principles of Population Genetics 4e*, Sinauer Associates (2007).
[Wei] Weinberg, W. Über den Nachweis der Vererbung beim Menschen, *Jahreshefte des Verein für vaterländische Naturkunde in Württemberg* 64:368–382 (1908).
[Yul] Yule, GU. Notes on the theory of association of attributes in statistics. *Biometrika* 2:121–134 (1903).