A Clarification of the Hardy-Weinberg Law

Alan E. Stark

(This research is self-funded – no institutional affiliation)

1 Address: 3/20 Seaview Street, Balgowlah, NSW, Australia 2093.

I thank Professors Francisco M. Salzano and Paulo A. Otto for their encouragement.
Running head: Clarification of Hardy-Weinberg Law

Key words: Hardy-Weinberg Law, non-random mating

Corresponding author: Dr. Alan E. Stark, 3/20 Seaview Street, Balgowlah NSW, Australia 2093

e-mail address: ae_stark@ihug.com.au

Phone: +61 2 99489038
ABSTRACT

C.C. Li showed that Hardy-Weinberg proportions (HWP) can be maintained in a large population by non-random mating as well as random mating. In particular he gave the mating matrix for the symmetric case in the most general form possible. Thus Li showed that, once HWP are attained, the same proportions can be maintained by what he called pseudo-random mating. This paper shows that, starting from any genotypic distribution at a single locus with two alleles, the same in each sex, HWP can be reached in one round of non-random mating with no change in allele frequency. In the model which demonstrates this fact, random mating is represented by a single point in a continuum of non-random possibilities.
INTRODUCTION

Li (1988) made a valuable contribution when he showed that Hardy-Weinberg (H-W) frequencies can be maintained in large populations with non-random mating. However, textbooks which invoke random mating as a justification for use of H-W frequencies in population models and further, maintain that random mating is a necessary condition, continue to be published. For example, Holsinger (2001) states “… the Hardy-Weinberg Law provides a way to estimate allele frequencies … provided we are willing to assume that all of the assumptions apply to the population in which we are interested”. The important point to notice here is that one of Holsinger’s assumptions is that individuals “choose mates at random”.

Random mating is sometimes described as pairing in ignorance of the genotypes of potential mates. But even lack of such knowledge does not ensure pairings with frequencies satisfying the formal criterion of randomness. Moreover, it is claimed frequently that assortative mating is incompatible with H-W frequencies. But this is not generally true. Random mating is not a realistic assumption for human population models and therefore it should be invoked only with careful qualification based on correct mathematical foundations. Because it is simple to describe, there has been too facile resort to the assumption of random mating. But this does not justify presentations of the most basic concept of population genetics which contain dubious and even formally incorrect notions.

Li (1988) wrote “An infinite number of patterns of deviations from random mating exists for autosomal loci that would make a population pseudo-random mating. This could be a contributing factor to the robustness of the Hardy-Weinberg law. … makes the study of the mating pattern a worthwhile subject.”
The next section gives the notation together with a brief outline of Li’s model, for the reader’s convenience. The main section gives the model which yields HWP from an arbitrary genotypic distribution in one round of non-random mating.

LI’s MODEL

Consider a population with respect to a single locus having alleles $A$ and $B$ with respective frequencies $q$ and $p$, the same in males and females. Denote frequencies of genotypes $AA$, $AB$ and $BB$ by $f_0$, $f_1$ and $f_2$. Table 1 gives Li’s symmetrical mating model which he introduces with the remark: “… When reciprocal crosses have the same frequency, the general pattern will be symmetrical.” Thus the roles of males and females can be reversed without changing the model. This case is simpler than Li’s more general model but is suitable for the present purpose. The $3 \times 3$ matrix of cell frequencies will be denoted by $[f_{ij}]$, $i = 0, 1, 2; j = 0, 1, 2$.

Both row totals and column totals give Hardy-Weinberg proportions: $f_0 = q^2$, $f_1 = 2pq$, $f_2 = p^2$. Thus the parental population is in H-W form and it is simple to show, under the usual assumptions, that the distribution of genotypes among offspring is the same. Note that $f_{11} = 4f_{02}$.

There are 4 parameters or constraints in Li’s model, taking up the four degrees of freedom in $[f_{ij}]$: these are $q$, Sewall Wright’s fixation index $F$, here taking the value zero, $a$ and $b$. Parameters $a$ and $b$ are constrained by the requirement that the elements of $[f_{ij}]$ be non-negative. In Li’s model, random mating is defined by the pair of conditions $a = 0$ and $b = 0$.

A compact formula giving a range of non-random mating tables which reproduce HWP is given by Stark (2005). This was adapted from a more general formula which
was used by Stark (1980) to classify systems of partial inbreeding. Other relevant references are Stark (1977a, b).

HWP FROM AN ARBITRARY DISTRIBUTION WITH NON-RANDOM MATING

This section demonstrates that H-W proportions can be obtained in one round of non-random, as well as random mating, from any genotypic distribution shared by males and females. Suppose that, in generation \( t \), the population has gene frequencies \( q_t \) and \( p_t \) and genotypic frequencies \( f_0(t) = q_t^2 + F_t p_t q_t \), \( f_1(t) = 2p_t q_t(1 - F_t) \) and \( f_2(t) = p_t^2 + F_t p_t q_t \). The possible values of \( F_t \) are constrained by \( q_t \) to ensure that \( \{f_0(t), f_1(t), f_2(t)\} \) is a valid set of genotypic frequencies, but generally \( F_t \) is in the interval \((-1, 1)\). Without loss of generality take \( q_t \) in the interval \( 0 < q_t \leq \frac{1}{2} \).

Consider a mating system in which the frequency of \( i \times j \) couples in generation \( t \) giving rise to offspring in generation \( t+1 \) is

\[
f_{ij}(t) = f_i(t)f_j(t)(1 + \mu d_i(t)d_j(t)/S_t + \nu e_i(t)e_j(t)/T_t), \quad (i = 0, 1, 2; \ j = 0, 1, 2), \ldots \ (1)
\]

where the entries in (1) are defined by

\[
d_0(t) = -2p_t, \quad d_1(t) = q_t - p_t, \quad d_2(t) = 2q_t.
\]

\[
S_t = 2p_t q_t(1 + F_t)
\]

\[
e_0(t) = p_t(F_t - 1)/(q_t + F_t p_t), \quad e_1(t) = 1, \quad e_2(t) = q_t(F_t - 1)/(p_t + F_t q_t),
\]

\[
T_t = p_t q_t(1 - F_t^2)/((q_t + F_t p_t)(p_t + F_t q_t)).
\]

Formula (1) was derived from Fisher’s Identity whose properties and background references are set out fully in Lancaster (1969, p. 90). The set of values \( \{d_0(t), d_1(t), d_2(t)\} \), denoted by \( d(t) \), was constructed by first assigning values 0, 1 and 2 respectively to genotypes \( AA \), \( AB \) and \( BB \), then correcting by deducting the mean with respect to the distribution of genotypic frequencies which is \( 2p_t \). Thus the mean of \( d(t) \), defined by \( \sum f_i(t)d_i(t) \), is zero and the variance \( S_t \) of \( d(t) \) is calculated from
The set of values \( \{ e_0(t), e_1(t), e_2(t) \} \) is denoted by \( e(t) \). Since the mean of \( e(t) \), defined by the expression \( \sum f_i e_i(t) \), is zero, the variance of \( e(t) \), denoted \( T_t \), is calculated from \( \sum f_i(e_i(t))^2 \). In (1) the quantity \( e_i(t)e_j(t)/T_t \) is the product of a pair of standardized elements, that is elements whose mean is zero and standard deviation unity.

This paper requires a special case of formula (1), namely the case when \( \mu = 0 \). Then (1) reduces to

\[
f_{ij}(t) = f_i(t)f_j(t)(1 + \nu e_i(t)e_j(t)/T_t), \quad (i = 0, 1, 2; \ j = 0, 1, 2). \quad \text{… (2)}
\]

Although equation (1) is not used in its entirety in what follows, (1) is included because otherwise (2) gives little insight to the reader as to its origins. Fisher’s Identity was conceived originally as a means of expressing a bivariate distribution in canonical form in order to explore its properties. For a \( 3 \times 3 \) matrix such as that in Table 1 this entails finding two sets of variable values each with mean zero and standard deviation unity. In this paper the procedure is reversed by starting from a canonical form, such as (1), but giving it desired properties. It turns out that \( d(t) \) as defined above, combined with setting the correlation coefficient \( \mu \) to be zero generates HWP in the offspring. Other values of \( \mu \) produce, in time, non-HWP.

Straightforward calculation shows that the distribution of genotypes in offspring of generation \( t+1 \) obtained from (2) is

\[
f_0(t+1) = q_t^2, \quad f_1(t+1) = 2p_tq_t, \quad f_2(t+1) = p_t^2,
\]

that is the offspring are distributed in HWP.

The admissible values of \( \nu \) in (2) are obtained by setting various elements of the mating matrix to zero. Given \( q_t \) and \( -q_t/p_t < F_i < 1 \), the upper limit is obtained as

\[
\nu_{up} = q_t(1 + F_i)(p_t + F_iq_t),
\]

when \( f_{01}(t) = 0 \).
The lower limit is set in two intervals: given $q_t$ and $-q_t/p_t < F_t \leq \frac{1}{2}(p_t - q_t)/p_t$, the left lower limit is obtained as

$$v_{ll} = (F_t + 1)q_t(q_t + F_t p_t)/((F_t - 1)p_t(p_t + F_t q_t)),$$

when $f_{00}(t) = 0$;

given $q_t$ and $\frac{1}{2}(p_t - q_t)/p_t \leq F_t < 1$, the right lower limit is obtained as

$$v_{rl} = p_t q_t(F_t^2 - 1)/((q_t + F_t p_t)(p_t + F_t q_t)),$$

when $f_{11}(t) = 0$.

As an example of (2), when the lower limits meet, that is when $f_{00}(t) = 0$ and $f_{11}(t) = 0$, $F_t = \frac{1}{2}(p_t - q_t)/p_t$, $e_0(t) = -1$, $e_1(t) = 1$, $e_2(t) = -q_t/(2 - 3q_t)$, $T_t = q_t(3 - 4q_t)/(2 - 3q_t)$, $v_{rl} = -T_t$, the mating matrix is as given by Table 2.

In summary, given any gene frequency $q$ and any admissible $F$, a value of $v$ can be chosen from an interval governed by $q$ and $F$, according to the limits given above, to produce HWP in one generation. Random mating is defined by the single value $v = 0$ contained in this interval. Thus Li’s contribution and this one completely remove the necessity of random mating as a requirement for the establishment and maintenance of HWP. Hardy and Weinberg showed that random mating was a sufficient condition and this is the concept that has taken hold in the genetics community for almost one hundred years.
LITERATURE CITED


Table 1.

Li’s symmetric non-random mating model

<table>
<thead>
<tr>
<th>♂ × ♀</th>
<th>AA</th>
<th>AB</th>
<th>BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$q^4 + a$</td>
<td>$2pq^3 - a - b$</td>
<td>$p^2q^2 + b$</td>
</tr>
<tr>
<td>AB</td>
<td>$2pq^3 - a - b$</td>
<td>$4p^2q^2 + 4b$</td>
<td>$2p^3q + a - 3b$</td>
</tr>
<tr>
<td>BB</td>
<td>$p^2q^2 + b$</td>
<td>$2p^3q + a - 3b$</td>
<td>$p^4 - a + 2b$</td>
</tr>
</tbody>
</table>
Table 2.

Mating scheme with $f_{00}(t) = 0$ and $f_{11}(t) = 0$

<table>
<thead>
<tr>
<th>♂ × ♀</th>
<th>AA</th>
<th>AB</th>
<th>BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>$q_t^2$</td>
<td>$\frac{1}{2}q_t(1-2q_t)$</td>
</tr>
<tr>
<td>AB</td>
<td>$q_t^2$</td>
<td>0</td>
<td>$p_t q_t$</td>
</tr>
<tr>
<td>BB</td>
<td>$\frac{1}{2}q_t(1-2q_t)$</td>
<td>$p_t q_t$</td>
<td>$1-3q_t+2q_t^2$</td>
</tr>
</tbody>
</table>