

Estimation of Divergence from Hardy–Weinberg Form

Alan E. Stark

School of Mathematics and Statistics, University of Sydney, New South Wales, Australia

The Hardy–Weinberg (HW) principle explains how random mating (RM) can produce and maintain a population in equilibrium, that is, with constant genotypic proportions. When proportions diverge from HW form, it is of interest to estimate the fixation index F , which reflects the degree of divergence. Starting from a sample of genotypic counts, a mixed procedure gives first the orthodox estimate of gene frequency q and then a Bayesian estimate of F , based on a credible prior distribution of F , which is described here.

■ **Keywords:** Hardy–Weinberg equilibrium, non-random mating, fixation index, Bayesian estimation

The HW distribution, usually called Hardy–Weinberg equilibrium (HWE),

$$\{q^2, 2q(1 - q), (1 - q)^2\}, \quad (1)$$

is widely used in population genetics as a base or platform for development of theory and analysis (Edwards, 2008; Hartl & Jones, 2006; Mayo, 2008; Penrose, 1972; Russell, 2006). In (1), q is the frequency (proportion) of one of two alleles in the population. Stark and Seneta (2014) have given a more general model, referred to as general mating equilibrium (GME), which does not assume RM. There are many examples in the genetics literature where genotypic proportions close to HW proportions have been observed. When subject to statistical test, they are found to be not significantly different from HW proportions. In other cases, it may be of interest to estimate the degree of divergence from HW proportions, which is measured by an index F , sometimes called the *fixation index*, at other times the *coefficient of inbreeding*. The GME model is partly characterized by the marginal distribution of F , which is described below. Here, given a sample of genotypic counts $\{n_{UU}, n_{UT}, n_{TT}\}$, I suggest that this distribution is a credible prior distribution of F that can be employed to yield the posterior distribution of F , a procedure called *Bayesian*, since it uses a formula introduced by Bayes (1763).

The mating system is defined completely by the matrix C , as described in the next section. Because the specification of the limits of its elements requires some detail, a geometrical description is used. However, this is simply a device to simplify the presentation.

Much of the appeal of the HW model, as published by Hardy (1908) and Weinberg (1908), is due to its apparent simplicity. Weinberg (1909), in his review of the first edition of Johannsen's monumental book published in the same year, states that Johannsen erred in attributing the model to Hardy, since Weinberg said it was due to [Karl] Pearson and, following Pearson, to himself. Stern (1943) brought Weinberg's paper to the attention of the genetics community generally. Crow (1999) explains the slowness of recognition of Weinberg's contribution to genetic analysis in the English-speaking world.

Li (1988) showed that it is possible to have HWE with non-random mating (NRM), hence Li's phrase 'pseudorandom mating'. The same possibility is implicit in a formula of Stark (1980). Stark (2006a) showed that it is possible to reach HW form with one round of NRM. It is still possible to find statements that RM is a necessary condition for HW proportions — for example, on page 664 of Russell (2006), though this is by no means rare. Weinberg (1908) used the HW distribution to analyze the inheritance of twinning in man. Stark (2006b) noted that Weinberg's analysis may need to be modified in the light of Li's paper. Stark and Seneta (2012) show that a paper of a Russian mathematician S. N. Bernstein was a fundamental, though largely unrecognized, contribution to genetic analysis by reason of its connection with the HW law.

RECEIVED 29 April 2015; ACCEPTED 1 June 2015. First published online 3 July 2015.

ADDRESS FOR CORRESPONDENCE: Dr Alan E. Stark, Honorary Associate, School of Mathematics and Statistics, F07, University of Sydney, NSW 2006, Australia. E-mail: alans@exemail.com.au

The GME model is more general than HWE, since it applies to a range of values of F and so includes HWE ($F=0$) as a special case.

Schull (1965) has a section on the estimation of the coefficient of inbreeding which he defines as:

$$F = \sum \left(\left(\frac{1}{2} \right)^{n_s+n_d+1} (1 + F_A) \right),$$

where F_A is the coefficient of inbreeding of any common ancestor that makes the connecting link between a line of ancestry tracing back from the sire and one tracing back from the dam. The numbers of generations from sire and dam to such a common ancestor are designated n_s and n_d respectively. Schull says that this concept may be extended to a population average coefficient of inbreeding. This interpretation may be applied to F as used in this paper. Schull describes practical and theoretical difficulties in estimating F , presumably based partly on experiences reported in Schull et al. (1962). In particular, Schull gives an example of the instability of some orthodox methods of estimating F . Schull et al. (1962) estimated F to be about 0.006 in their study. Neel and Schull (1954, p. 73) give the following formula for the frequency with which recessively inherited conditions arise from consanguineous marriages:

$$Fq + (1 - F)q^2.$$

If the gene frequency of the deleterious gene is $q=1/1,000$ and $F=0.006$, the condition is about seven times more frequent in the inbred population than in a population in which $F=0$.

The need to estimate F arises in forensic investigations. Balding and Nichols (1995) give a formula which contains F . In explaining use of the formula, they say that F is analogous to F_{ST} of Sewall Wright; that is, it concerns stratification of populations in which gene frequencies vary from stratum to stratum. Ayres and Balding (1998) use symbol f , 'a parameter measuring departure from HW caused by inbreeding'. They discuss a number of difficulties associated with estimating f , including choice of prior distributions of f , but do not consider the choice of a *logical prior* such as that given here. They give a composite posterior distribution of f calculated from a sample of Samoans resident in New Zealand, which has a mode at about $f=0.05$.

Cavalli-Sforza and Bodmer (1971, pp. 377–379) have a short section on estimating F , having in mind deviation from HWE due to inbreeding. This method would be classified as orthodox and so is outside the scope of this article.

In the next section, for convenience, I summarize the essential features of GME. The following section gives the marginal distribution of F , then the Bayesian estimation procedure with the aid of an example, and finally a brief discussion mainly about Fisher's (1959) view of the appropriate use of Bayes' formula.

The General Mating Equilibrium Model

We deal only with a single autosomal locus with two alleles U and T with frequencies in the population q and p ($q + p = 1$). Throughout, q remains constant because this is guaranteed by the nature of the selected mating system. A set of frequencies of genotypes $\{UU, UT, TT\}$ can be represented in terms of q and a measure of departure from HW form F as, say, $\mathbf{a}' = \{q^2 + Fpq, 2pq - 2Fpq, p^2 + Fpq\}$. These will vary according to F and will be denoted generally by $\{f_0, f_1, f_2\}$, ($f_0 + f_1 + f_2 = 1$), that is $f_0 = q^2 + Fpq$, and so on.

The population is maintained in discrete generations according to the mating scheme:

$$\begin{bmatrix} UU \times UU & UU \times UT & UU \times TT \\ UT \times UU & UT \times UT & UT \times TT \\ TT \times UU & TT \times UT & TT \times TT \end{bmatrix} \quad (2)$$

with commensurate pairing frequencies given by the matrix

$$C = \begin{bmatrix} f_{00} & f_{01} & f_{02} \\ f_{10} & f_{11} & f_{12} \\ f_{20} & f_{21} & f_{22} \end{bmatrix}. \quad (3)$$

C is symmetric, that is $f_{ij} = f_{ji}$, with row and column sums $\{f_0, f_1, f_2\}$. This triple of sums is the parental frequency distribution.

Below we use C in the extended (row vector) form

$$\mathbf{u}' = \{f_{00}, f_{01}, f_{02}, f_{10}, f_{11}, f_{12}, f_{20}, f_{21}, f_{22}\}. \quad (4)$$

To follow the progression of generations, we need Mendel's coefficients of heredity given in matrix form by

$$M = \begin{bmatrix} 1 & 1/2 & 0 & 1/2 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1 & 1/2 & 1/2 & 1/2 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 1/2 & 1 \end{bmatrix}. \quad (5)$$

Then, the frequency distribution of juveniles is calculated from

$$\mathbf{j}' = (M\mathbf{u})'$$

which in detail is

$$\mathbf{j} = \left\{ f_{00} + \frac{f_{01} + f_{10}}{2} + \frac{f_{11}}{4}, \frac{f_{01}}{2} + f_{02} + \frac{f_{10} + f_{11} + f_{12}}{2} + f_{20} + \frac{f_{21}}{2}, \frac{f_{11}}{4} + \frac{f_{12} + f_{21}}{2} + f_{22} \right\}'. \quad (6)$$

The population is in equilibrium, that is, the distribution of juveniles is the same as that of adults, if and only if matrix C has, in addition to the properties given above, the special property

$$f_{11} = 4f_{02} = 4f_{20}. \quad (7)$$

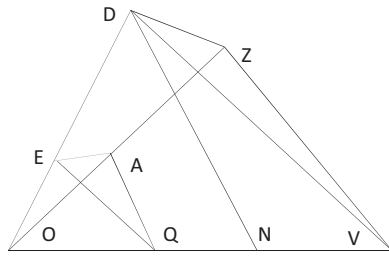


FIGURE 1
Schematic illustration of the bounding region of admissible sets of F , f_{11} , and f_{01} , for $1/4 < q < 1/2$.
Note: The admissible region is defined by the vertices Q , V , Z , D , E , A . The region defined by vertices O , Q , A , and E is not part of the admissible region. The coordinates of the vertices are given in Table 1. Coordinates of points of reference not shown on Table 1 are: $O (-q/p, 0, 0)$; $B ((p - 2q)/(3p), 0, 0)$; $N ((2p - q)/(3p), 0, 0)$.

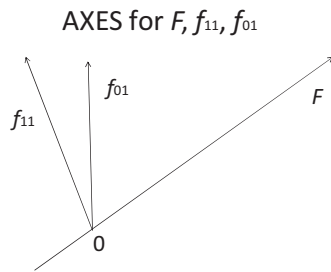


FIGURE 2
Orthogonal axes used to specify coordinates F , f_{11} , and f_{01} for given q .

The notation used here is a modified version of that given in Stark and Seneta (2013, 2014).

By way of explanation, special condition (7), on the system of mate choice, is that the frequency of mating pairs heterozygote \times heterozygote is four times the frequency of each of the reciprocal mating pairs homozygote \times opposite homozygote. This applies to HWE under RM where the first kind of mating has frequency $(2pq)^2$ and each of the second kind $(q^2) \times (p^2)$. Under this condition, the ‘loss’ of heterozygotes from the first kind of mating is exactly compensated by the ‘gain’ from the second kind of mating. Identity (7) allows for NRM as well as RM.

Matrix C is a complete description of the mating system. However, the elements of C are subject to various constraints that are conveniently delineated by a geometrical figure that consists of two tetrahedrons, one of which is contained within the other, as shown schematically in Figure 1. For the reader’s convenience, the characteristics of this figure are spelled out in more detail below using schematics from Stark and Seneta (2014).

I assume first that q is fixed so C is defined by the trio (F, f_{11}, f_{01}) , which can be depicted by points in Euclidean space using orthogonal coordinate axes as shown in Figure 2. The regions of admissible points are of three main

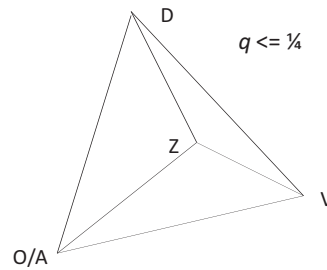


FIGURE 3
Schematic illustration of the bounding region of admissible sets of F , f_{11} , and f_{01} for $q \leq 1/4$; vertex O replaces A when $q < 1/4$.

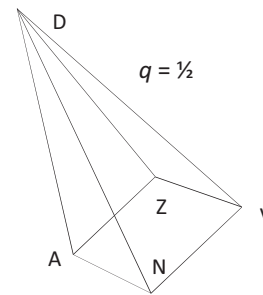


FIGURE 4
Schematic illustration of the bounding region of admissible sets of F , f_{11} , and f_{01} for $q = 1/2$.

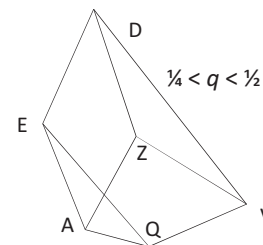


FIGURE 5
Schematic illustration of the bounding region of admissible sets of F , f_{11} , and f_{01} for $1/4 < q < 1/2$.

types, depending on q : (1) $1/4 < q < 1/2$; (2) $q \leq 1/4$; (3) $q = 1/2$. The faces of the regions are planar, the planes defined by $f_{11} = 0$ (Λ_1) and $f_{01} = 0$ (Λ_2) being two main ones. The remaining planes are defined by the following equations:

$$\Lambda_3 \quad 4pqF - f_{11} - 4f_{01} + 4q^2 = 0$$

$$\Lambda_4 \quad 2pqF + f_{11} + f_{01} - 2pq = 0$$

$$\Lambda_5 \quad 12pqF + 3f_{11} + 4f_{01} + 4p(p - 2q) = 0.$$

Only planes Λ_1 – Λ_4 are relevant when $q < 1/4$. The respective admissible regions are shown schematically in Figures 3, 4, and 5.

The Marginal Distribution of F

Case $q \leq 1/4$

This section describes properties of the model depicted in Figure 3 of Stark and Seneta, 2014 — reproduced here as Figure 3. Suppose that it defines a trivariate distribution of variables F, f_{11} , and f_{01} , which maintain a genotypic distribution $\{f_0 = q^2 + Fpq, f_1 = 2pq - 2Fpq, f_2 = p^2 + Fpq\}$. Suppose further that points are distributed uniformly within the space enclosed by the bounding planes. For fixed $q \leq 1/4$, the volume enclosed is $S = 4q^2/(27p)$. This can be demonstrated by taking slices of the solid around values of f_{11} , giving triangles with base $(4q - 3f_{11})/(4pq)$, ($p = 1 - q$), and height $(4q - 3f_{11})/6$. The area of a triangular section is $(16q^2 - 24qf_{11} + 9f_{11}^2)/(48pq)$. Integrating this quantity with respect to f_{11} over the range 0 to $4q/3$, which are the limits of f_{11} , gives the volume.

Suppose that a population with a particular gene frequency q arrives randomly at some point within the defined space. The value of F will be governed by the marginal distribution of F , here denoted by $\Pi(F)$. This function has the following values:

$$\Pi_1(F) = 2f_0^2/S, \quad -q/p \leq F \leq (p - 2q)/(3p) \quad (8)$$

$$\Pi_2(F) = 2[q^2 + (f_1 - f_0)(4f_0 - f_1)]/(9S), \quad (p - 2q)/(3p) < F \leq (2p - q)/(3p) \quad (9)$$

$$\Pi_3(F) = f_1^2/(2S), \quad (2p - q)/(3p) < F \leq 1. \quad (10)$$

The distribution of F , $\Pi(F)$, has mode $(p - q)/(2p)$. This can be demonstrated by differentiating $\Pi_2(F)$ with respect to F and equating the derivative to zero. The mode is midway between $(p - 2q)/(3p)$ and $(2p - q)/(3p)$, consistent with the fact that $\Pi(F)$ is symmetrical about the mode. The distribution at the mode is $f_0 = q/2, f_1 = q, f_2 = (2p - q)/2$.

The symmetry of $\Pi_2(F)$ can be demonstrated by calculating the value of $(f_1 - f_0)(4f_0 - f_1)$ at points equidistant from modal F . For a point above the mode, take $F^+ = (p - q + \Delta)/(2p)$, ($\Delta > 0$), and the corresponding one below $F^- = (p - q - \Delta)/(2p)$. Taking first F^- , $f_1 - f_0 = q(1 + 3\Delta)/2$, and $4f_0 - f_1 = q(1 - 3\Delta)$, so $(f_1 - f_0)(4f_0 - f_1) = q^2(1 - 9\Delta^2)/2$. In the case of F^+ , $f_1 - f_0 = q(1 - 3\Delta)/2$, and $4f_0 - f_1 = q(1 + 3\Delta)$ and $(f_1 - f_0)(4f_0 - f_1) = q^2(1 - 9\Delta^2)/2$.

The value of $\Pi_1(F)$, namely $2f_0^2$, is equivalent to the area of a right triangle with height $f_{01} = f_0$ and base $f_{11} = 4f_0$, so that when $F = F^-$, $2f_0^2 = q^2(1 - \Delta)^2/2$. Similarly, the value of $\Pi_3(F)$, namely $f_1^2/2$, is equivalent to the area of a right triangle with base $f_{11} = f_1$ and height $f_{01} = f_1$, so that when $F = F^+, f_1^2/2 = q^2(1 + \Delta)^2/2$. Thus, symmetry of $\Pi(F)$ extends over the whole interval $(-q/p, 1)$.

The distribution of F , $\Pi(F)$ has modal value $9/5$ when $F = 1/5$. This can be demonstrated by differentiating $\Pi_2(F)$ with respect to F and equating the derivative to zero.

TABLE 1
The Coordinates of the Vertices of the Admissible Regions as Functions of q

Vertex	F	f_{11}	f_{01}
A	$-(p - q)^2/4pq$	0	$q - 1/4$
V	1	0	0
D	$(p - 2q)/(3p)$	$4q/3$	0
E	$-(1 - 4q + 6q^2)/(6pq)$	$2(4q - 1)/3$	0
Z	$(2p - q)/(3p)$	0	$2q/3$
Q	$(3q - 1)/(3q)$	0	0

Case $q = 1/2$

This section describes properties of the model depicted in Figure 4 of Stark and Seneta (2014) — reproduced here as Figure 4. For fixed $q = 1/2$, the volume enclosed is $S = 1/27$. The marginal distribution of F , here denoted by $\underline{\Pi}(F)$, has the following values:

$$\underline{\Pi}_1(F) = (1 + 3F)^2/(24S), \quad F \leq 0 \quad (11)$$

$$\underline{\Pi}_2(F) = (1 + 6F - 15F^2)/(24S), \quad 0 < F \leq 1/3 \quad (12)$$

$$\underline{\Pi}_3(F) = f_1^2/(2S), \quad 1/3 < F \leq 1. \quad (13)$$

Differentiating $\underline{\Pi}_2(F)$ with respect to F and equating to zero shows that the mode of $\underline{\Pi}(F)$ is $1/5$.

Case $1/4 < q < 1/2$

This section describes properties of the model depicted in Figure 2 of Stark and Seneta (2014) for $1/4 \leq q \leq 1/2$ — reproduced here as Figure 5. The volume of the solid is

$$S = \frac{16q^3 - (4q - 1)^3}{108pq}. \quad (14)$$

Depending on the value of q , the distribution of F is defined using elements of the following set of functions of q and F :

$$h_1 = (1 - 4q + 6f_0^2)/(3S) \quad (15)$$

$$h_2 = (2(1 - 4q + 6f_0^2) - (4f_0 - f_1)^2)/(6S) \quad (16)$$

$$h_3 = 2(f_0)^2/S \quad (17)$$

$$h_4 = 2(3f_0f_1 - q^2)/(3S) \quad (18)$$

$$h_5 = (f_1)^2/(2S) \quad (19)$$

$$h_6 = 2(3(f_0)^2 - (f_1 - f_2)^2)/(3S) \quad (20)$$

$$h_7 = (8f_1(1 - f_1) - (4(f_0)^2 + 5(f_1)^2 + 4(f_2)^2))/(6S). \quad (21)$$

Figure 6 shows the marginal distribution of F for $q = 38/111$. The above functions are needed because the relative order of various defining points change with changes of q . The distribution of F is defined according to Table 2. For

TABLE 2
Definition of the Marginal Distribution of F for $1/4 < q < 1/2$

$q, F h$	F	h	F	h	F	h	F	h	F	h
$q \rightarrow$	1/4–1/3	1/4–1/3	1/3	1/3	1/3–3/8	1/3–3/8	3/8	3/8	3/8–1/2	3/8–1/2
F	E–A	1	E–A	1	E–A	1	E–A	1	E–B	1
F	A–Q	6	A–Q	6	A–B	6	A–Q	7	B–A	2
F	Q–B	3	—	—	B–Q	7	—	—	A–Q	7
F	B–N	4	Q–N	4	Q–N	4	Q–N	4	Q–N	4
F	N–V	5	N–V	5	N–V	5	N–V	5	N–V	5

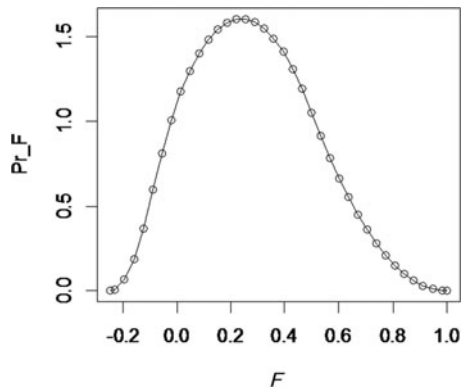


FIGURE 6
The marginal distribution of F for $q = 38/111$ (Pr_F stands for probability density of F).

example, when $q = 1/3$, the function is

$$\Pi(F) = h_1, h_6, h_4, h_5 \tag{22}$$

over the respective intervals of F : E–A, A–Q, Q–N, and N–V.

The distribution $\Pi(F)$ of F has mode $(1 - 2q)/(2(1 - q))$. This gives the distribution of genotypes:

$$\{f_0 = q/2, f_1 = q, f_2 = 1 - 3q/2\}. \tag{23}$$

The Estimation of F

Cavalli-Sforza and Bodmer (1971, p. 43) give data relating to the MN blood-group locus: 47 M, 52 MN, and 12 N individuals. There are 76 N genes from a total 222 and so the frequency of gene N is $q = 38/111$. These are the counts used to illustrate the estimation method. The method of gene counting is the orthodox method of estimating q . I treat this as the ‘true’ value of q although clearly the value 38/111 is subject to sampling error. The standard error of estimate is available.

Figure 6 displays the (marginal) distribution of F using $q = 38/111$ which I employ as the prior distribution to get the posterior distribution of F shown in Figure 7.

If the value of the fixation index is F , the genotypic proportions in the population are as follows:

$$\{f_0 = q^2 + Fpq, f_1 = 2pq - 2Fpq, f_2 = p^2 + Fpq\}. \tag{24}$$

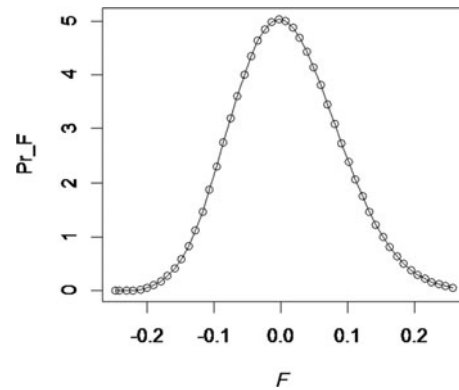


FIGURE 7
The posterior distribution of F computed from genotypic counts {12, 52, 47} from which $q = 38/111$ (Pr_F stands for probability density of F).

Denoting the genotypic counts by $\{n_{UU}, n_{UT}, n_{TT}\}$, the (conditional) probability of observing these counts is as follows:

$$C(F) = \frac{n!}{n_{UU}! \times n_{UT}! \times n_{TT}!} \times f_0^{n_{UU}} \times f_1^{n_{UT}} \times f_2^{n_{TT}}, \tag{25}$$

where n is the sample size.

If $P(F).dF$ is the prior probability that F lies in an infinitesimal interval containing F , then, following Thomas Bayes, the posterior probability that it is in that interval is as follows:

$$P'(F).dF = \frac{P(F).dF \times C(F)}{\int P(F) \times C(F).dF}. \tag{26}$$

The posterior distribution of F from (26) for the above counts is displayed in Figure 7.

Discussion

Fisher (1959) begins Chapter II as follows:

For the first serious attempt known to us to give a rational account of the process of scientific inference as a means of understanding the real world, in the sense in which this term is understood by experimental investigators, we must look back over two hundred years to an English clergyman, the Reverend Thomas

Bayes, whose life spanned the first half of the eighteenth century.

In that chapter, Fisher touches on the views of various eminent thinkers such as Gauss, Laplace, Montmort, de Moivre, Boole, and Jeffreys. Most of the text dwells on the problem of deciding whether an acceptable prior distribution is available; in particular, about when is there a suitable axiomatic prior. On this last point, on page 18, Fisher makes the comment [about an axiomatic prior]: ‘... the question, more natural to an experimental investigator, of whether, in the particular circumstances of the investigation, the knowledge implied by the postulate was or was not in fact available.’

On pages 18–20, Fisher gives an example of what he regards as a valid use of Bayes’ formula. It relates to determining the genotype of a black mouse mated to a brown mouse. The critical point is whether there is knowledge of the origin of the mouse being tested.

Scientific inference remains a contentious subject. Fisher (1960, p. 197), using the phrase *inverse probability* for the application of Bayes’ formula, writes: ‘Statements of inverse probability ... require for their truth the postulation of knowledge beyond that obtained by direct observation.’ Good (1968, p. 7) states: ‘I regard it as mentally healthy to believe that credibilities exist ...’. He says that a *credibility* is ‘a rational intensity of conviction, implicit in the given information, and such that if a person does not agree with it he is wrong’. An alternative term is *logical probability*. I regard the method described here as using a *logical* prior probability distribution.

Smith (1959), an advocate for Bayesian methods, writes as follows:

When Bayes’ Theorem can be applied, it is more informative than a significance test, for it gives to each hypothesis an exact probability of being true. A significance test, on the other hand, may ‘reject a hypothesis at significance level P ’, but P here is not the probability that the hypothesis is true, and indeed the rejected hypothesis may still be probably true if the odds are sufficiently in its favour at the start. For example, in human genetics there are odds of the order of 22:1 in favour of two genes chosen at random being on different chromosomes; so even if a test indicates departure from independent segregation at the 5 per cent level of significance, this is not very strong evidence in favour of linkage.

Although his monograph is concerned mainly with classical (orthodox) statistical methods, Weir (1996) has a section on Bayesian methods, including an example due to Gunel and Wearden (1995) and another due to Lange (1995). These use empirical, rather than logical, prior distributions. Weir gives classical methodology for estimating F .

Acknowledgment

I thank the referee for suggesting ways to improve the manuscript.

References

- Ayres, K. L., & Balding, D. J. (1998). Measuring departures from Hardy-Weinberg: A Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity*, 80, 769–777.
- Balding, D. J., & Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, 96, 3–12.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370–418.
- Cavalli-Sforza, L. L., & Bodmer, W. F. (1971). *The genetics of human populations*. San Francisco: W. H. Freeman and Company.
- Crow, J. F. (1999). Hardy, Weinberg and language impediments. *Genetics*, 152, 821–825.
- Edwards, A. W. F. (2008). G. H. Hardy (1908) and Hardy-Weinberg equilibrium. *Genetics*, 179, 1143–1150.
- Fisher, R. A. (1959). *Statistical methods and scientific inference* (2nd ed.). Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1960). *The design of experiments* (7th ed.). Edinburgh: Oliver and Boyd.
- Good, I. J. (1968). *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge, MA: The MIT Press.
- Gunel, E., & Wearden, S. (1995). Bayesian estimation and testing of gene frequencies. *Theoretical and Applied Genetics*, 91, 534–543.
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28, 49–50.
- Hartl, D. L., & Jones, E. W. (2006). *Essential genetics: A genomics perspective* (4th ed.). Sudbury, MA: Jones and Bartlett Publishers.
- Lange, K. (1995). Applications of the Dirichlet distribution to forensic match probabilities. *Genetica*, 96, 107–117.
- Li, C. C. (1988). Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics*, 119, 731–737.
- Mayo, O. (2008). A century of Hardy-Weinberg equilibrium. *Twin Research and Human Genetics*, 11, 249–246.
- Neel, J. V., & Schull, W. J. (1954). *Human heredity*. Chicago: The University of Chicago Press.
- Penrose, L. S. (1972). *The biology of mental defect* (4th ed.). London: Sidgwick & Jackson.
- Russell, P. J. (2006). *iGenetics: A molecular approach* (2nd ed.). San Francisco, CA: Pearson Education.
- Schull, W. J. (1965). Estimation of genetic parameters in population studies. In J. V. Neel, M. W. Shaw, & W. J. Schull (Eds.), *Genetics and the epidemiology of chronic diseases* (pp. 45–60). Washington, DC: U. S. Department of Health, Education, and Welfare.

- Schull, W. J., Yanase, T., & Nemoto, H. (1962). Kuroshima: The impact of religion on an island's genetic heritage. *Human Biology*, 34, 271–298.
- Smith, C. A. B. (1959). Some comments on the statistical methods used in linkage investigations. *American Journal of Human Genetics*, 11, 289–304.
- Stark, A. E. (1980). Inbreeding systems: Classification by a canonical form. *Journal of Mathematical Biology*, 10, 305.
- Stark, A. E. (2006a). A clarification of the Hardy-Weinberg law. *Genetics*, 174, 1695–1697.
- Stark, A. E. (2006b). Stages in the evolution of the Hardy-Weinberg law. *Genetics and Molecular Biology*, 29, 589–594.
- Stark, A. E., & Seneta, E. (2012). On S. N. Bernstein's derivation of Mendel's law and 'rediscovery' of the Hardy-Weinberg distribution. *Genetics and Molecular Biology*, 35, 388–394.
- Stark, A. E., & Seneta, E. (2013). A reality check on Hardy-Weinberg. *Twin Research and Human Genetics*, 16, 782–789.
- Stark, A. E., & Seneta, E. (2014). Hardy-Weinberg equilibrium as foundational. *International Journal of Statistics in Medical Research*, 3, 198–202.
- Stern, C. (1943). The Hardy-Weinberg law. *Science*, 97, 137–138.
- Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahreshefte d. Vereins f. vaterländische Naturkunde in Württemberg*, 64, 369–382.
- Weinberg, W. (1909). Johannsen, W. Elemente der exakten Erblchkeitslehre. Deutsche wesentlich erweiterte Ausgabe. IV. u. 516 S. JenaFischer [Review]. *Archiv. für Rassen-und Gesellschafts-Biologie*, 6, 553–555.
- Weir, B. S. (1996). *Genetic data analysis II*. Sunderland, MA: Sinauer Associates.
-