IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 1 Ver. I (Jan. - Feb. 2017), PP 01-03
www.iosrjournals.org

Considerations on the genetic equilibrium law

Simone Camosso
(Università degli studi di Milano Bicocca, Italy)

Abstract: *In the first part of the paper I willpresentabriefreview on the Hardy-Weinberg equilibrium and it's formulation in projective algebraicgeometry. In the second and last part I willdiscussexamples and generalizations on the topic.*
Keywords:*Hardy-Weinberg, likelihood function, maximum likelihood estimation.*

I. Introduction

The study of population genetics, evolution and its evolutionary trees are classical subjects in biology. A mathematical approach consists in the maximum likelihood estimation, a technique largely used in statistic. This approach leads to the problem of maximizing particular functions of certain parameters. A theoretical study and an upper bound for the maximum likelihood degree is discussed in [3]. Different techniques in order to solve likelihood equations are described in [6]. Applications of these methods have been used when the statistical model is an algebraic variety, this is the case of Fermat hypersurfaces treated in [1]. In the other side we have applications of these ideas to biological models. In this direction we refer to a work of [4] where phylogenetic models in two different topologies have been studied by the authors. The purpose of this paper is a "soft" introduction to these ideas with a discussion on the Hardy-Weinberg case.

II. The Hardy-Weinberg law

The Hardy-Weinberg law states that allele and genotype frequencies in a population remain constant during the generation change. This happens under the following assumptions: the size of the population must be very large, we have absence of migration and mutations, the mating is random and the natural selection doesn't affect the alleles under consideration. Mathematically if p represents the number or pure dominants characters *AA*, q the number of heterozygotes *Aa* and r the number of pure recessives *aa*, the following proportion holds $p{:}2q{:}r$ (see [5]). Another way, if p and q represent the allele frequencies of the character *A* and*a* with $p+q=1$, taking the square we find that:

$$p^2 + 2pq + q^2 = 1, \tag{2.1}$$

where p^2, $2pq$ and q^2represent the genotype frequencies associated to *AA,Aa* and *aa*. The equation (2.1) describes the constancy of the genotypic composition of the population and is called the Hardy-Weinberg principle or the Hardy-Weinberg equilibrium (HWE). We consider [11] and [12] as scholarly references on this subject. Different generalizations of (2.1) are possible. The first concerns the number of alleles at a locus. For example in the case of three alleles A_1, A_2 and A_3, with frequencies respectively given by p,q and r, the genotype frequencies are given by the following expansion $(p+q+r)^2=p^2+q^2+r^2+2pq+2qr+2pr$. In general, for any number n of alleles with frequencies p_i, we have that:

$$(p_1+\ldots+p_n)^2 = \sum_{i=1}^{n} p_i^2+\sum_{i\neq j}^{n} p_i\, p_j\,. \tag{2.2}$$

In another direction the generalization is given considering the binomial $(p+q)^m$, with $m=3,4,5,\ldots$. This is the case of polyploid. For example considering tetraploids ($m=4$) the procedure involves the expansion of $(p+q)^4$. We observe how in this particular example the frequency of heterozygotes (given by the mixed terms in the expansion) is $2pq(2-pq)$ that is considerably greater then $2pq$, the frequency for a diploid organism. More information on this topic can be found in [11].

III. Projective and algebraic geometry

In this section we shall examine how translate previous considerations in the modern language of projective and algebraic geometry. The setting is the same of [8] in its first lecture. We shall show how the Hardy-Weinberg law can be formulated in a fixed system of homogeneous coordinates l,m,nin P^2. First we shall consider the open triangle $\Delta_2(l,m,n) = \{R_+^3 :\ l+m+n=1\}$, where $R_+$are the positive reals. Second we shall observe that setting $l=p^2,m=2pq,n=q^2$ to be genotype frequencies, we have the relation:

$$m^2 = 4nl, \tag{3.1}$$

that is the equation of a parabola in the triangle of vertex *(1,0,0),(0,1,0),(0,0,1).* We call the zero locus of (3.1) denoted also by $V(m^2-4ln)$, the Hardy-Weinberg curve (details are in [10]). In the theory of [8] (and [9]) is of particular interest a function called "likelihood function" that depends by some positive integer parameters. This function is positive on $\Delta_2$and zero on the boundary ofΔ_2. We shall denote this function by *l* and with $u_0,u_1,\ldots$ the corresponding parameters. In the case of the Hardy-Weinberg curve this function has the following form:

$$l_{u_0,u_1,u_2} = l^{u_0}\, m^{u_1}\, n^{u_2} = 2^{u_1}\, p^{2u_0+u_1}\, q^{u_1+2u_2}.$$

We observe that $l_{u_0,u_1,u_2}$is a function depending only by the variable *p* (because *q=1-p*) and the MLE problem consists in the estimation of *p* maximizing the function l_{u_0,u_1,u_2}.Lagrange Multipliers can be used in order to solve the problem and in this case the solution is given by the point:

$$\hat{p} = \frac{2u_0 + u_1}{2u_0 + 2u_1 + 2u_2}.$$

IV. Examples, generalizations and conclusion

As exercise we shall apply the same procedure in order to solve the MLE problem for the case of three alleles and in the second time for the case of tetraploids. For the first we shall start writing the "likelihood function" associated to the HWE given by $(p+q+r)^2$, where *p,q* and *r* are the usual frequencies. In this case we have that:

$$l_{u_0,u_1,u_2,u_3,u_4,u_5} = 2^{u_3+u_4+u_5}\, p^{2u_0+u_3+u_5}\, q^{2u_1+u_3+u_4}\, (1-p-q)^{2u_2+u_4+u_5}.$$

We shall proceed maximizing the function of two variables *p,q*. This is an ordinary problem of calculus that gives as answer the point:

$$\hat{p} = \left(\frac{2u_0+u_3+u_5}{2u_1+2u_2+2u_4+u_3+u_5},\ \frac{u_4-u_5+2u_1-2u_0}{2u_1+2u_2+2u_4+u_3+u_5},\ \frac{2u_2+u_5+u_4}{2u_1+2u_2+2u_4+u_3+u_5}\right).$$

For the tetraploid case, before to proceed, we shall observe that calling$l_0=p^4,l_1=q^4,l_2=4pq^3,l_3=4p^3q,l_4=6p^2q^2$the genotype frequencies, the Hardy-Weinberg equilibrium can be represented by the following relation:

$$l_4^4 = \frac{1}{81} l_0 l_1 l_2 l_3.$$

The associated "likelihood function" is

$$l_{u_0,u_1,u_2,u_3,u_4} = l_0^{u_0}\, l_1^{u_1} l_2^{u_2} l_3^{u_3} l_4^{u_4}.$$

We shall make the expedient of consider the logarithm of the previous function instead the original finding as maximizing point:

$$\left(\frac{u_0}{|u|},\ \frac{u_1}{|u|},\ \frac{u_2}{|u|},\ \frac{u_3}{|u|},\ \frac{u_4}{|u|}\right),$$

where $|u| = u_0 + u_1 + u_2 + u_3 + u_4$ .We recommended the use of a scientific software, as Maple or MATLAB, especially when the number of parameters is considerably high. Now I want to spend these last words comparing analogies between the HWE and the algebraic geometry. It is clear that a possible extension of the Hardy-Weinberg law can take the following form:

$$(p_1+\ldots+p_n)^m = c, \tag{4.1}$$

where*c* is some constant and p_i from *i=0,…,n* are the allele frequencies such that the sum is fixed. Now expanding (4.1) we find the polynomial form:

$$\sum_{i_0+\cdots+i_n=m} \frac{m!}{i_0!\ldots i_n!} p_0^{i_0} \ldots p_n^{i_n} = c.$$

From the algebraic geometry point of view this is the image of the Veronese map$v_m: P^n \to P^{\binom{n+m}{m}-1}$ given by $(p_0,\ldots,p_n) \mapsto (p_0^m, p_0^{m-1}p_1, \ldots, p_n^m)$ (see [2]). The classical Hardy-Weinberg law corresponds to the case of $v_1: P^1 \to P^2$ that $(p,q) \mapsto (p^2, pq, q^2)$ and $c=1$. Using the identification between homogeneous polynomials that are power of linear forms and the image of the Veronese map, we can think these generalized laws as Veronese projective varieties. From the side of algebraic geometry there are a rich collection of results concerning the Veronese and Segre varieties, for example it is possible to compute the Hilbert polynomial and other invariants. It is not all peace and light because the constraint $\Delta_n = \{(p_0,\ldots,p_n) \in R_+^{n+1} : p_0 \ldots + p_n = 1\}$ doesn't permit the complete translation of the problem using the previousidentification.

Anyway the methods of numerical algebraic geometry seem to give good prospects in this direction and in [7] the ML degree has been calculated for matrices with rank constraints. In particular the case of rank one gives the ML degree equal to one, so $\hat{p}$ is a rational function of a set of parameters $u_0, u_1, \ldots$.

References

Journal Papers:

[1]. D.Agostini, D.Alberelli, F.Grande, P.Lella, The maximum likelihood degree of Fermat hypersurfaces, arXiv:1404.5745.
[2]. E.Arrondo, Introduction to projective varieties, unpublished notes from the website: http://www.mat.ucm.es/~arrondo/projvar.pdf (2007), 9-10.
[3]. F.Catanese, S.Hoşten, A.Khetan, B.Sturmfels, The maximum likelihood degree, Amer. J. Math. 128 (2006), no. 3, 671-697. MR 2230921.
[4]. B.Chor, A.Khetan, S.Snir, Maximum likelihood on four taxa phylogenetic trees: analytic solutions, The 7th Annual Conference on Research in Computational Molecular Biology-RECOMB 2003, Berlin, April 2003, pp. 76-83.
[5]. G.H.Hardy, Mendelian Proportions in a Mixed Population, Science, New Series, Vol.28, 706 (1908), 49-50.
[6]. S.Hoşten, A.Khetan, B.Sturmfels, Solving the Likelihood Equations, Foundations of Computational Mathematics, Vol. 5, Issue 4, pp 389-407, 2005.
[7]. J.Hauenstein, J.Rodriguez, B.Sturmfels, Maximum Likelihood for Matrices with Rank Constraints, Journal of Algebraic Statistics, Vol.5, Issue 1 (2014), pp 18-38.
[8]. J.Huh, B.Sturmfels, Likelihood Geometry, Combinatorial Algebraic Geometry: LevicoTerme, Italy 2013, Springer International Publishing, Vol.2108 of the series Lecture Notes in Mathematics (2014), 63-117.
[9]. I.J.Myung, Tutorial on maximum likelihood estimation, Journal of Mathematical Psychology 47 (2003) 90-100.

Books:

[10]. A.W.F.Edwards, Foundations of Mathematical Genetics, Cambridge University Press, Cambridge (2000).
[11]. R.Frankham, J.D.Ballou, D.A.Briscoe, Introduction to conservation genetics, Cambridge (2002), 86-90.
[12]. D.L.Hartl, A.G.Clark, Principles of population genetics, Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts (Fourth Edition).