Battling about equations

One of the most exciting episodes in the study of human heredity was the battle between two incompatibe algebraic expressions to describe its properties in the beginning of the 20th Century. In one corner was Karl Pearson (1857–1936) with his 1898 formulation for ancestral heredity, which had the authority of Charles Darwin and Francis Galton for support. It reads:

\[ f(y) = (0.5)\sigma_0/\sigma_1 k_1 + (0.25)\sigma_0/\sigma_2 k_2 + (0.125)\sigma_0/\sigma_3 k_3 + (0.062)\sigma_0/\sigma_4 k_4 \ldots \]

where \( \sigma_0/\sigma_n \) are the ratios of the standard deviations of measurable traits in the offspring to the standard deviations of the mid-parental generation; and \( k_1, k_2 \ldots k_n \) are the deviations of the mid-parental means from the mean of the offspring for each trait. He admits to reservations of how ‘mid-parent’ should be defined. The subscripts refer to each of the ancestral generations (parents, grandparents, great-grandparents, etc.). The original paper should be consulted for fuller details.2

Pearson elaborated on earlier formulations using a convergent geometric series to represent the hereditary contribution \( f(y) \) by the parents, the four grand parents, eight great-grand parents, etc. to the offspring.1 This type of convergent series is of the general form:

\[ f(y) = A + Bx + Cx^2 + Dx^3 + \ldots \]

where \( A, B, C, D \ldots \) are constants independent of \( x = 0.5^n \) in Pearson's expression. The expression can be used to represent a great many biological phenomena. The formulation has no theoretical significance; all it postulates is that the phenomenon in question varies continuously. Then Maclaurin’s or Taylor’s theorems (depending on the number of variables involved) can be used to determine the values of the coefficients that will make the series useful to any desired degree of approximation.3

Pearson’s equation certainly satisfies the requirements for continuous variation of inherited traits which Darwin believed to be true; and it reaches a limit of one with no inflection points (if \( \sigma_0 / \sigma_n \) and \( k_n < 1 \)). Pearson was very self-congratulatory about his equation, covertly comparing himself to Isaac Newton; which is surprising in view of its resemblance to a Taylor series. He wrote at the end of his paper:2 ‘If Darwinian evolution be natural selection combined with heredity then the single statement which embraces the whole field of heredity must prove as epoch-making to the biologist as the law of gravitation to the astronomer.’

In the opposite corner was William Bateson (1861–1926) who had adopted Mendel’s formulation4 by 1902. For the F2 generation the equation can be represented as:

\[ (A + a)(A + a) = AA + 2Aa + aa, \]

where \( A \) and \( a \) are parental alleles for random segregation to gametes, and then randomly combined in the offspring, (in accordance with Mendel’s Laws of Inheritance). It is astonishing that such a simple algebraic expression can accurately describe the actual phenomena of inheritance during reproduction. The left hand part of the equation represents the hybridization of the two parental genotypes; the right hand part of the equation is the assortment of alleles (assumed to be random) found as genotypes in the offspring.

This equation resembles an expansion of the binomial theorem, i.e. \((A+a)^2\); but it does not follow the ordinary rules of algebra since on the right hand side of the equation (as found in Mendel’s original paper4) AA does not signify \( A^2 \); biologically it signifies an additive function of alleles (i.e. \( A+A \)); there appears to be no multiplicative properties of alleles. Mendel4 represented it as \( A/A \) which would be confusing terminology nowadays, as this indicates the operation of division. (However indices do have a place in the Hardy–Weinberg expression because the probability of combined events is multiplicative,
e.g. $p^2 + 2pq + q^2$, where (p) and (q) now represent the allele frequencies in a large, randomly mating population, not undergoing selection, from which one can calculate the frequencies of expected genotypes. The right hand part of Mendel’s equation can also be considered as a probability distribution where $P(\text{AA}) + P(\text{not AA}) = 1$. In Mendel’s formulation biological weighting is given to the alleles so that (A) is a dominant allele and (a) is a recessive allele as found individually in the male and female gametes. For such a binomial-type distribution the possible outcome of events in the offspring of a single pedigree can be classified as either positive or negative, that is to say whether a specific trait (e.g. black urine) either will occur or not. It predicts that the combination of (aa) would occur on average once in every four sibs; that 2 sibs would be Aa, and one sib would be AA, and one with aa, hence giving a probability for aa of $P = 0.25$. This was the average distribution found by Garrod in 1902 in his study of the inheritance of the recessive black urine trait of alkaptonuria in first cousin marriages.\(^5\) Garrod played no further part in the dispute because he appears to have been a rather modest and self-effacing man.\(^6\)

Bateson’s adoption and promotion of Mendel was anathema to Pearson who was enamoured by his own formulation. Pearson, as editor of the journal *Biometrika* of which he was a co-founder, would not allow any of Bateson’s papers on Mendel’s ideas to be published in it. The battle therefore developed between the Biometricians, led by Pearson, and the Mendelians, led by Bateson. Pearson was intolerant of any dissent from his own views. For Pearson inheritance was purely a matter of continuous graded variation for natural selection to work on as formulated in his equation; for Bateson it was a matter of discrete discontinuous events that can follow a binomial probability distribution to produce distinct variations in body structures due to the inheritance of particulate units as deduced from Mendel’s experiments with the garden pea.

Pearson’s formulation is a kind of solution and probably does represent an approximate expression for the heredity of quantitative characters but is not a particularly useful one. It is too general a statement and could be used, like a Taylor series, to model many other things. Even though the formulation is intuitively plausible it is of little practical use because it fails to predict anything. No amount of clever mathematics like this could lead anyone to deduce that inherited factors come in pairs (one of each pair from each parent) or that one of the factors can be dominant over the other in its biological expression. On the other hand Mendel’s formulation fits many experimental results and predicts that a particular recessive character will appear (under certain defined conditions) in the next generation with a ratio of 1:4 siblings. This makes all the difference, because it points to a discoverable underlying mechanism giving rise to this probability ratio.

The battle came to a head at the Cambridge British Association meeting in 1904.\(^7\) There was a fierce showdown between the two schools of thought. The Pearson camp used the two usual arguments against Bateson and the Mendelians: (i) there were often intermediate features found in many hybrid crosses that suggest blending inheritance as proposed by Darwin, and (ii) the Mendelian results take no account of ancestral inheritance and their data could be explained by other models than the one proposed by Mendel. Moreover Bateson’s methods were to be condemned as careless and his theories about underlying mechanisms as ‘cumbersome and undemonstrable’. Bateson responded to this after the lunch break. He flatly rejected the criticism and stated ‘soon every science that deals with animals and plants will be teeming with discoveries made possible by Mendel’s work’. He then likened Pearson’s camp to ‘flat-earthers’ who had resolutely described the paths of the heavenly bodies to harmonize with a theory of the flatness of the earth; just as the Pearson group were harmonizing the facts of heredity with false ideas about blending inheritance. The chairman of the meeting, the Reverend T R Stebbing, a self-appointed ‘man of peace’ after suggesting that compromise is a good thing went on: ‘You have all heard what Professor Pearson has suggested (…about a possible armistice…), but what I say is: let them fight it out.’ The succeeding battle became intense; there was to be no compromise. The Mendelians stated categorically that inheritance must be based on discrete particles; whereas the Pearson camp maintained that it was statistically based on the blending of inherited properties. The controversy became quite bitter. Bateson was later to write to his wife Beatrice of one of his opponents (Walter Weldon): ‘If any man ever set himself to destroy another man’s work that did he do to me…’.\(^8\)

The battle was eventually won by the Mendelians who have held centre stage ever since. By 1911, Pearson had still not given up: ‘there is no definite proof of Mendelism applying to any living form at present’.\(^9\) Pearson was a great and original man but very intransigent in his opinions. However, Mendelian analysis breaks down when considering the inheritance of common metabolic disorders such as Type II diabetes mellitus, atherosclerosis or the dyslipidaemias. Despite knowing many of the
genetic components of say, hypertriglyceridaemia, the clinical outcome can still only be predicted within a large range of probabilities. There is an urgent need for expressions to describe the rules of the various types of non-Mendelian inheritance (polygenic, epigenetic, gene conversion, etc.). These will need to combine all the newly gained genetic information to give a much greater certainty to prediction of clinical outcome. In many cases, an appropriate expression might well turn out to be in the form of a convergent geometric series, as Pearson used. At present it appears to be done by calculating the cumulative dosage of risk alleles for each incriminated locus, but with no means of biological weighting,$^{10}$ as the concepts of ‘dominance’ and ‘recessive’ have done for Mendelian analysis.

David J. Galton

Acknowledgements

Grateful thanks are due to Dr John Harris, Professor Mike Wallis and Professor John Beardmore for helpful criticisms.

References