

although it is not necessary to evaluate all of these, the amount of calculation makes anything but a computer analysis impracticable. A program was therefore written to carry out an exact test of significance. This test gives a composite probability for the pooled samples of 0.0002, indicating that the results are highly significant. Fisher's approximation, on the other hand, gives a probability of between 0.6 and 0.5 ($\chi^2 = 26.6$; 28 d.f.).

Although these samples are undoubtedly extreme examples of discrete distributions, such a gross difference in the results of the two tests illustrates the dangers involved in the indiscriminate application of Fisher's method.

R. H. McLAUGHLIN

School of Zoology,
University of New South Wales,
Kensington, Australia.

Received October 2, 1968.

¹ Fisher, R. A., *Statistical Methods for Research Workers*, thirteenth ed. revised (Oliver and Boyd, Edinburgh, 1958).

² Wallis, W. A., *Econometrica*, **10**, 229 (1942).

McLAUGHLIN gives an interesting illustration of the serious error which may arise from the incorrect application to discrete distributions of Fisher's method of combining significance probabilities. Wallis's exact computer solution is one that is more readily available nowadays; but it seems also worth noting an approximate solution applicable to the present context.

Let the cumulative probabilities of the possible samples in the *i*th experiment, listed in order of their significance in relation to the alternative hypothesis, be

$$P_{ir} = \sum_{j=1}^r p_{ij},$$

and let the actual sample have $r = s$. Compute $x_{ir} = -\log_e P_{ir}$, $m_i = \sum_r p_{ir} x_{ir}$, $\sigma_i^2 = \sum_r p_{ir} (x_{ir} - m_i)^2$, and take $z = \sum_i (x_{is} - m_i) / \sqrt{\sum_i \sigma_i^2}$ as an approximate standardized normal variate. In the example given (of course, on the assumption that the P_{ir} have all been recorded, for otherwise the m_i and σ_i could not be correctly calculated), this seems to give a significance probability of 0.000025, corresponding to $z = 4.06$. This rather exaggerates the exact value 0.0002 noted by McLaughlin, but is vastly less misleading than the χ^2 value. One obvious source of inaccuracy is the skewness of z ; this could be computed and corrected for, but while this scarcely seems worth while, it does seem useful to make a rough correction obtained from the theoretical skewness in the continuous case. This yields the amended value

$$z' = z - \frac{1}{3} (z^2 - 1) / \sqrt{n}$$

where n is the number of tests, or 2.68 for the example, with corresponding significance level 0.004, an over-correction but still indicating high significance.

M. S. BARTLETT

Department of Biomathematics,
University of Oxford.

Received October 21, 1968.

Fundamental Theorem of Natural Selection

FISHER'S fundamental theorem of natural selection has recently been discussed^{1,2}. Edwards² observes that the theorem $V' - V \geq 0$, where V and V' represent, in successive discrete generations, the mean fitness at a single diallelic locus in a random mating diploid population, has already been proved by Moran³. He notes further that the

more general result involving a multiple allelic system was actually established some years ago, and comments that the reason for this curiosity is that the validity of the theorem for two alleles was simply assumed to follow immediately from the work of Sewall Wright.

In fact, Scheuer and I⁴ have indicated that if a, b, h, p and q are non-negative numbers with $p + q = 1$, and if

$$Vp' = p(ap + hq), \quad Vq' = q(hp + bq) \quad (1)$$

with

$$V = ap^2 + 2h pq + bq^2 \quad (2)$$

and

$$V' = ap'^2 + 2hp'q' + bq'^2 \quad (3)$$

then it can be shown by means of elementary but somewhat heavy algebra that

$$V' - V = \frac{1}{V^2} \{p(ap + hq - V)^2 + q(hp + bq - V)^2\} \\ \{V + ap + bq\} \quad (4)$$

from which the theorem in question immediately follows.

Moreover, it can easily be shown that equation (4) is equivalent to the formulation given by Li⁵. For if, following Li, we set

$$\beta = (a - h)p + (h - b)q \quad (5)$$

then, with a little algebra, we find that

$$ap + hq - V = \beta q, \quad hp + bq - V = -\beta p \quad (6)$$

and

$$ap + bq = V + (a - 2h + b)pq \quad (7)$$

Thus equation (4) readily becomes

$$V' - V = \frac{pq\beta^2}{V^2} \{2V + (a - 2h + b)pq\} \\ = \frac{2pq\beta^2}{V} + (a - 2h + b) \left\{ \frac{pq\beta^2}{V} \right\}^2 \quad (8)$$

which is, apart from the change in notation, precisely equation (7) of Li.

Furthermore, it follows from equation (1) and some algebra that the change in gene frequency in one generation can be written in the forms

$$p' - p = \frac{\beta pq}{V} \quad (9)$$

and

$$p' - p = \frac{pq}{V} \{(h - a) + (h - b)\} \{P - p\} \quad (10)$$

where

$$P = \frac{h - b}{(h - a) + (h - b)} \quad (11)$$

is clearly the equilibrium gene frequency if such in fact exists.

Equation (9) is in fact identical with the expression given by Li for the change in gene frequency; in the form of equation (10), however, it enables the other question raised by Edwards to be settled, namely, that of the monotonic convergence of the gene frequency. In fact, in either of the situations $a < h < b$ or $a > h > b$ in which one allele has an unconditional selective advantage over the other, β is respectively negative and positive for all values of p and q and it follows that the gene frequency will change monotonically in the direction of extinction of the inferior allele. In the case $h > a, h > b$ where the heterozygote is better adapted than either homozygote, we have clearly

$$0 < \frac{pq}{V} \{(h - a) + (h - b)\} < 1 \quad (12)$$

and it follows from equations (10) and (12) that the gene frequency p' always lies between the values of p and P ; that is to say, the gene frequency is approaching the equilibrium monotonically, as Edwards suggests.