

## NOTES AND COMMENTS

### THE NUMBER TO BREED TO BE REASONABLY CERTAIN OF BEING ABLE TO REJECT ONE OF TWO HYPOTHESES

A. W. F. EDWARDS

*Istituto di Genetica, University of Pavia, Italy*

Received 5.i.62

A simple problem, familiar to genetics students, is that of determining how many offspring should be bred in order to be reasonably certain of being able to distinguish between expected  $s : 1-s$  and  $t : 1-t$  segregations. The usual technique is to determine the number required to make it possible for both hypotheses to be on the limit of rejection at a given level of significance. This somewhat arbitrary criterion is good enough for purposes of experimental design: if one of the hypotheses is indeed true, and the recommended number is bred, then in the great majority of cases the other hypothesis will be able to be rejected with confidence. This note treats the general case of  $n$ -class segregations.

Suppose there are  $Na_i$  individuals in the  $i$ th class, and the alternative hypotheses are  $E(Na_i) = Ns_i$  and  $E(Na_i) = Nt_i$ , where  $\sum a_i = \sum s_i = \sum t_i = 1$ ,  $N$  being the total number. All sums are taken over the  $n$  classes. Then twice the log likelihood ratios on the two hypotheses are

$$G_s = 2\sum Na_i (\log Na_i - \log Ns_i) \text{ and } G_t = 2\sum Na_i (\log Na_i - \log Nt_i).$$

For the case in which both hypotheses are on the limit of rejection,  $G_s = G_t = \chi_{n-1}^2$ ,  $\chi_{n-1}^2$  being that value appropriate to the chosen level of significance. We may write this equation because  $G$  is approximately equal to  $\chi^2$  (see Woolf, 1957). From it we find

$$\sum a_i \log s_i = \sum a_i \log t_i \tag{1}$$

and 
$$\sum a_i (\log a_i - \log s_i) = \chi_{n-1}^2 / 2N. \tag{2}$$

We already have 
$$\sum a_i = 1. \tag{3}$$

These three equations cannot by themselves lead to an explicit solution for the  $a_i$  and  $N$  for  $n > 2$ . In order to obtain a solution we must choose the "worst possible" values of the  $a_i$ , and hence the maximum value of  $N$ , by minimising the left-hand side of equation (2) subject to the conditions (1) and (3). We must therefore minimise the function

$$H = \sum a_i (\log a_i - \log s_i) + \rho (\sum a_i \log s_i - \sum a_i \log t_i) - \sigma (\sum a_i - 1),$$

in which  $\rho$  and  $\sigma$  are Lagrangian multipliers. We find

$$\frac{\partial H}{\partial a_i} = \log a_i - \log s_i + 1 + \rho (\log s_i - \log t_i) - \sigma = 0,$$

$$\frac{\partial^2 H}{\partial a_i^2} = 1/a_i > 0 \text{ and } \frac{\partial^2 H}{\partial a_i \partial a_j} = 0.$$

We now have  $n+2$  equations to determine the  $n$   $a_i$ 's,  $\rho$  and  $\sigma$ , and hence  $N$ .

Fortunately we may eliminate all the  $a_i$ 's, whose values we do not want to know, as follows. Multiplying each equation by  $a_i$  and summing over all  $i$  gives  $\sum a_i (\log a_i - \log s_i) + (1-\sigma)\sum a_i + \rho \sum a_i (\log s_i - \log t_i) = 0$ ,

whence, using equations (1), (2) and (3),  $\sigma - 1 = \chi_{n-1}^2/2N$ . Further, each equation may be written  $a_i = \sigma' t_i^p / e s_i^{p-1}$ , where  $\sigma = \log \sigma'$ . Summing over all  $i$  gives

$$\chi_{n-1}^2/2N = \sigma - 1 = -\log e/\sigma' = -\log \sum t_i^p / s_i^{p-1}. \quad (4)$$

Finally, substituting for  $a_i$  in (1),

$$\sum t_i^p / s_i^{p-1} \log t_i / s_i = 0. \quad (5)$$

### APPLICATION

In many genetical situations there will only be a few classes, with the result that equation (5) may be solved for  $p$  without difficulty. The substitution of this value in (4) then leads to the required number  $N$ . In some cases it will be possible to derive an expression for  $N$  directly.

For example, consider the case in which two loci are known to be linked with recombination fraction  $p = 1 - q$ ; we wish to test whether a double heterozygote has the dominant alleles in coupling or repulsion. On backcrossing to the double recessive, the segregation of the four phenotypes amongst the progeny will be either  $p/2 : q/2 : q/2 : p/2$  or  $q/2 : p/2 : p/2 : q/2$ . Substitution of these values in (5) shows that, independently of  $p$ ,  $p = \frac{1}{2}$ , and substituting in (4) then gives  $\chi_{3}^2/2N = -\log 2\sqrt{pq}$ , whence

$$N = -\chi_{3}^2 / \log 4pq.$$

### REFERENCE

- WOOLF, B. 1957. The log likelihood ratio test (the G-test). *Ann. Hum. Genet.*, Lond., 21, 397-409.  
 Note added in proof: see BAILEY. 1961. *Introduction to the Mathematical Theory of Genetic Linkage*, Oxford: Clarendon Press, p. 30; for the case of two classes and a recommendation that the  $\chi^2$  value should be "one-tailed".

## POPULATION STRUCTURE, BREEDING SYSTEM, INTER-SPECIFIC HYBRIDISATION AND ALLOPOLYPLOIDY

T. P. PALMER

*Crop Research Division, Department of Scientific and Industrial Research,  
Christchurch, New Zealand*

### 1. INTRODUCTION

Received 14.xi.61

Inter-specific hybridisation in higher plants has three common results when sexual reproduction is retained in the hybrid population. A hybrid swarm may be produced; the hybrids may back-cross to the parent species, and so their hybrid origin may be obscured; or they may produce allopolyploid derivatives reproductively isolated from the parent species. Introgression is more likely to occur in cross-fertilisers than in self-fertilisers (Baker, 1953), while Grant (1956) has argued that since the union of un-reduced gametes is more likely to occur in rare self-fertilising hybrids than in rare cross-fertilising hybrids, allopolyploids are more likely to arise in self-fertilisers, and cites the high coincidence of polyploidy and self-fertility in some