NOTES AND COMMENTS

ON THE ESTIMATION OF THE NUMBER OF SEX ALLELES IN HONEYBEES

GUNNAR EKBOHM* and KERSTIN EBBERSTEN[†] *Division of Statistics and [†]Division of Bee Breeding, Swedish University of Agricultural Sciences, S-750 07 Uppsala

Received 15.ix.78

In honeybees sex is determined by a series of allelic genes at the sex-locus. Queens and workers are heterozygous at the sex-locus, drones are haploid or homozygous (diploid drones) but these homozygotes are lethal because the larvae are moved away by the workers leaving holes in the brood. Laidlaw *et al.* (1956) estimated the number of sex alleles in a population of bees to be about 12. In this note we discuss their statistical method and give some alternatives.

The material used in the Laidlaw *et al.* (1956) experiment was 65 hives of a random mating population which was in connection with about 200 other hives in the surroundings. Sixty-one virgin queens from these 65 hives were inseminated with just one drone each from one colony of the same population. Brood viability tests were made revealing the number of queens with low viable brood. Low viable brood means that only about 50 per cent (or less) of the brood was alive, depending on the degree of "inbreeding" and thus the 50 per cent diploid drones which die. The queen and the drone she mated with had one sex allele in common and about half of the progeny was homozygous. If the queen and the drone had different sex alleles, the brood had nearly 100 per cent viability.

In the sequel let *n* be the (unknown) number of sex alleles. It can be shown that if there are *n* alleles in the population equilibrium is attained when all alleles have the same frequency (1/n).

In analysing the experimental outcome Laidlaw *et al.* first consider the maximum likelihood estimator $\hbar = 2N/H$, where N is the number of colonies out of which H show low viability. However, since this is not an unbiased estimator—in fact it has not even a finite expectation—they reject it and suggest another estimator with rather a small bias,

$$n^* = 2(N+1)/(H+1).$$

An approximation of the variance of n^* is also derived. The result of the experiment was that out of 61 hives nine showed low viability and, hence, Laidlaw *et al.* estimate the number of alleles to be 12.4 with a variance of 12.69. If we assume n^* to be approximately normally distributed a 95 per cent confidence interval is found to be 12.4 ± 6.98 or, let us say 5.4 to 19.4. (This is, of course, a rather doubtful assumption but it is made here for comparison.)

If the number of sex alleles is *n* the expected proportion of colonies with low viability is equal to $\pi = 2/n$. Now, $(N(H/N - \pi)^2/[\pi(1 - \pi)]$ is

approximately distributed as χ^2 with one degree of freedom. In order to find confidence limits we set this variable equal to the $(1-\alpha)$ 100 per cent point of the χ^2 (1)-distribution. Then we have the following second degree equation

$$(\chi_{\alpha}^{2}+N)\pi^{2}-(\chi_{\alpha}^{2}+2H)\pi+H^{2}/N=0.$$

Solving this equation and setting π equal to 2/n the following limits for n are obtained

$$4(\chi_{\alpha}^{2}+N)/\{2H+\chi_{\alpha}^{2}(1\pm\sqrt{1+4H(N-H)}/N\chi_{\alpha}^{2})\}.$$

With the data given by Laidlaw *et al.* we then get the 95 per cent confidence limits to be 7.8 and 25.1.

Edwards (1972) states that in scientific inference the general problem is to assess the relative merits of rival hypotheses. In the present case the rival hypotheses are different values of the number of sex alleles. Then the relative merits are given by what Edwards calls the support curve, which is defined as the natural logarithm of the likelihood function. The maximum of this support curve is conventionally set equal to zero and is reached for the maximum likelihood estimator. For other values of the parameter, the number of sex alleles, the support relative to that value is then given by the support curve. Thus, it is possible to identify two equally supported parametric values between which all values have a larger support. These values are called m-unit support limits if all hypotheses within the interval have a support greater than m units less than the maximum.

The likelihood function in the present case is

$$L = \binom{N}{H} \pi^{H} (1-\pi)^{N-H} \quad \text{where } \pi = 2/n$$

The support curve is then obtained as

$$S(n) = (N-H) \ln \{(n-2)/(N-H)\} - N \ln (n/N) - H \ln (H/2)$$

For the data of Laidlaw *et al.* we find the 2-unit limits to be 7.9 and 27.6 and the 1-unit limits to be 9.1 and 22.0.

Laidlaw *et al.* (1956) is a reference commonly used to support the statement that the number of sex alleles in the honeybee is 12 (see *e.g.* Woyke, 1976). However, as shown above, their experimental result is not enough to give a very precise estimate and it also indicates that the number may be essentially larger. The more sex alleles there are in a population the easier it is to plan the breeding, due to the fact that the expected frequency of homozygotes decreases when the number of alleles increases. For a brief discussion of the consequences for bee breeding see Ebbersten (1976).

Often in statistics unbiased estimators are striven for and presumably for good reasons. But in a case like the one treated in this paper we think it is more correct to regard as the best estimate the value which is the one most supported by the experimental data, and this estimator is the maximum likelihood one. The shape of the support curve, *e.g.* in terms of support limits, is also of interest in getting knowledge about the true parameter value. In fact, Edwards (1972) finds the maximum likelihood point estimator unsatisfactory and that what is needed is a relative measure of support for rival hypotheses. Such a support curve he gets from the likelihood function. For a discussion of the method of support the reader is referred, e.g. to Barnett (1973).

References

BARNETT, v. 1973. Comparative Statistical Inference. Wiley, London.

- EBBERSTEN, K. 1976. Skattning av antalet könsalleler hos bina med hjälp av en modifierad beräkningsmetod ur ett försök utfört av Laidlaw *et al.* 1956. Nordisk Bitidskrift, 27, 49-54.
- EDWARDS, A. W. F. 1972. Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference. Cambridge University Press, London.
- LAIDLAW, H. H., GOMES., F. P., AND KERR, W. E. 1956. Estimation of the number of lethal alleles in a panmitic population of Apis Mellifera L. Genetics, 41, 179-188.
- WOYKE, J. 1976. Population genetic studies on sex alleles in the honeybee using the example of the Kangaroo Island bee sanctuary. *Journal of Apicultural Research*, 15, 105-123.