THE MEASUREMENT OF SEXUAL ISOLATION AND SELECTION: A CRITIQUE

J. M. GOUX† and D. ANXOLABEHRE††

† *Laboratoire de Génétique des Populations, Université Paris VII, Tour 42, 2 place Jussieu 75005 PARIS.*

Received 25.iii.80

SUMMARY

The theoretical basis of the sexual selection and isolation indices W and Z , often used by investigators in analysing population cage experiments, is reexamined. Their expectations and variances, under the null hypothesis of no selection and random mating, are a function of the number N of the matings and are also frequency-dependent. The exact probability of W is given for any value of N . It appears that the peculiar properties of the distribution of W make its use misleading.

1. INTRODUCTION

MANY theoretical models have been worked out which take into consideration the existence of sexual selection (O'Donald, 1977, 1978). But, it is necessary, when observing sexual associations at the population level, to distinguish the effects of isolation (homogamy) on one side and those of selection in each sex on the other. The problem then arises of providing a quantitative measure of these effects. In this paper, we consider only the analysis of experiments based on the mixture of both males and females of two different phenotypes A and B, the results of which can be presented in a 2×2 table (table 1). Several indices have been used by investigators to

TABLE 1

Number of mating pairs observed in a situation where males and females of two distinct types, A and B are present (no multiple matings)

Males		Females		
		A R_2	B $(N_2 - R_2)$	Σ N_2
A	R_1	a	b	$a + b$
B	$(N_1 - R_1)$	c	d	$c + d$
Σ	N_1	$a + c$	$b + d$	N matings

obtain a quantitative measure of sexual isolation or selection based on these kinds of results. We intend to consider principally three of them:

$$W_I = \frac{ad}{bc}, \quad \text{a measure of isolation,}$$

$$W_M = \frac{ab}{cd}, \quad \text{a measure of male sexual selection,}$$

$$W_F = \frac{ac}{bd}, \quad \text{a measure of female sexual selection.}$$

†† E.R.A. 406 du C.N.R.S.: "Analyse et mécanismes de maintien du polymorphisme".

Some other derived parameters, $Z_I = \sqrt{W_I}$, $Z_M = \sqrt{W_M}$, $Z_F = \sqrt{W_F}$ and $Y_I = (Z_I - 1)/(Z_I + 1)$ are equally in use. The coefficient K of Petit (1951) is only a special case of Z_M , when no influence of females is present, and is still used by some investigators (White and Grant, 1977; Ribo, 1977). The $W_I = ad/bc$ has also been used for the estimation of genetic linkage (Mather, 1951).

The theory of W is due to H. Levene and reported by Petit and Ehrman (1968) and by Ehrman and Petit (1968). Since these publications, many investigators have used these parameters, usually in experiments with *Drosophila* species (Elens, Van den Haute and Delcour, 1974; Petit, Kitagawa and Takamura, 1976; Fontdevila and Mendez, 1979).

In general, a number of N mating pairs is observed in a cage during a short period of time after mixing two types, A and B, in order to impede any multiple copulations. Thus, R_1 males A, $(N_1 - R_1)$ males B, R_2 females A, $(N_2 - R_2)$ females B are put together in the cage. The elements of the probabilistic model of Levene are given in table 2. According to Levene's

TABLE 2

Probabilities of the different types of matings under the null hypothesis of random mating and no sexual selection

		Females	
		A	B
Males	A	$P_1 = \frac{R_1}{N_1}$	$P_2 = \frac{R_2}{N_2}$
	B	$Q_1 = \frac{N_1 - R_1}{N_1}$	$Q_2 = \frac{N_2 - R_2}{N_2}$
		$p = P_1 P_2$	$p' = P_1 Q_2$
		$p'' = Q_1 P_2$	$p''' = Q_1 Q_2$

analysis, the observed value of each W parameter, W_I for instance, is compared to its expectation $E(W_I) = 1$ on the basis of a confidence interval using an estimation of the variance of W_I :

$$\text{Var}(W_I) = W_I^2 \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right).$$

The same is done for Z , but the correct variance of Z_I is:

$$\text{Var}(Z_I) = \frac{1}{4} Z_I^2 \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)$$

which differs from the incorrect formula given by Ehrman and Petit (1968) and Petit and Ehrman (1968).

2. THE UNDERLYING HYPOTHESES OF LEVENE'S MODEL

The use of these formulae, on the basis of the underlying probabilistic model, rests on several hypotheses:

Hypothesis 1: The probability of drawing an A (rather than a B) individual in each sex is binomially distributed under the null hypothesis of

no sexual selection. It means that the number N of mating pairs observed must be small compared to the number N_1 of males (resp. N_2 of females), in order for the marginal probabilities to be constant. We note in passing that this hypothesis is accepted in the UMPU test of association advocated by Kendall and Stuart, II, 33.24 (1973).

Hypothesis 2: The probability of drawing the N mating pairs is multinomial, the probability of each class of mating pairs being the product of the binomial probabilities of each type of mate, under the null hypothesis of random mating. Then, this distribution rests on the constancy of the marginal probabilities.

Hypothesis 3: The expectation of W_I is $E(W_I) = 1$. But it is easily shown that this is only verified when $N = \infty$.

Hypothesis 4: The use of a confidence interval $\pm \sigma_{W_I}$ to determine the significance of an observed W_I rests on the supposed symmetry, and even the normality, of the distribution of W_I .

We shall examine subsequently the validity of these hypotheses.

3. THE VALIDITY OF THESE HYPOTHESES

Hypothesis 1: In the particular conditions of these types of experiments, the number of individuals introduced in the observation cage cannot be high, for the different types of matings to be scored. As Spiess and Schwer (1978) have observed, the choice of an A (as opposed to a B) individual modifies the marginal probabilities. Then, the distribution of this drawing is not binomial, but hypergeometric.

Hypothesis 2: If the probability of drawing of each mate corresponds to a hypergeometric distribution, then the probability of a given type of mating changes during the process of pair formation and thus the distribution of the mating pairs is not multinomial. It can be shown that the correct distribution of the N mating pairs (unfixed marginal totals) is, under the null hypothesis of no sexual selection and random mating:

$$P(a, b, c, d|N) = P_1 \times P_2 \times P_3 \quad (1)$$

where:

$$P_1 = \frac{R_1!(N_1 - R_1)!N!(N_1 - N)!}{(a + b)!(N - a - b)!(R_1 - a - b)!(N_1 - R_1 - N + a + b)!N_1!}$$

is the probability that $(a + b)$ males A have paired without male selection,

$$P_2 = \frac{R_2!(N_2 - R_2)!N!(N_2 - N)!}{(a + c)!(N - a - c)!(R_2 - a - c)!(N_2 - R_2 - N + a + c)!N_2!}$$

is the probability that $(a + c)$ females A have paired without female selection, and

$$P_3 = \frac{(a + b)!(N - a - b)!(a + c)!(N - a - c)!}{a!b!c!d!N!}$$

is the probability that the N pairings are at random (no homogamy) with marginal totals fixed: $(a + b)$ males A, $(a + c)$ females A, $(b + d)$ males B and $(c + d)$ females B.

The probability of drawing $N(a, b, c, d)$ pairs being known, it is possible to obtain from it the distribution of W_I . But we have to consider the fact that the drawings of males (as opposed to females) are hypergeometric, on the basis of a finite value of N . But if N is finite, it means that a zero value for the frequency of a certain class of mating pairs has a non null probability. It follows that some drawings of $N(a, b, c, d)$ pairs can give indeterminate ($W_I = 0/0$) or infinite ($W_I = ad/0$) values. We note that this is also the case for the binomial model, as soon as, N becomes finite.

This creates a problem for the study of the distribution of W_I , because $E(W_I)$ and $\text{Var}(W_I)$ are not defined if W_I takes indeterminate values. Nevertheless, as we know $P(a, b, c, d/N)$ for every drawing, we can take an inventory of the drawings giving indeterminate W_I , and so obtain their total probability as well as the distribution of W_I , after discarding the indeterminate W_I values. $E(W_I)$ and $\text{Var}(W_I)$ are then defined, but they now take infinite values, since certain W_I values are infinite (Haldane, 1956). By eliminating these infinite values in the same way, we can finally obtain $P(W_I/\text{indeterminate or infinite } W_I \text{ excluded})$, as noted already by Edwards (1963).

Hypothesis 3: As N is finite, we cannot accept $E(W_I) = 1$ and we have to appreciate the bias $B_I = E(W_I) - 1$. At the same time, we shall question the meaning of the approximated variance $\text{Var}(W_I)$ given by Levene. It is easy to verify that Levene's formulae are obtained using the Taylor's series applied to $E(ad/bc)$ and $E(a^2d^2/b^2c^2)$ at the point $E(a), E(b), E(c), E(d)$, limiting the expansion to the first order only.

In the general case, where $E(ad/bc)$ and $\text{Var}(ad/bc)$ are indeterminate or take infinite values (for finite N), this expansion has no meaning. But the use of the truncated distribution obtained in the preceding paragraph permits us to give an approximate value to the bias B_I and a better approximation of the variance, if we expand to the second order, for instance.

By expanding to the second order we get:

$$E\left(\frac{ad}{bc}\right) = \frac{E(a)E(d)}{E(b)E(c)} \left[1 + \frac{\text{var } b}{E^2(b)} + \frac{\text{var } c}{E^2(c)} - \frac{\text{cov}(a, b)}{E(a)E(b)} - \frac{\text{cov}(a, c)}{E(a)E(c)} \right. \\ \left. + \frac{\text{cov}(a, d)}{E(a)E(d)} + \frac{\text{cov}(b, c)}{E(b)E(c)} - \frac{\text{cov}(b, d)}{E(b)E(d)} - \frac{\text{cov}(c, d)}{E(c)E(d)} \right] + \varepsilon$$

$$E\left(\frac{a^2d^2}{b^2c^2}\right) = \frac{E^2(a)E^2(d)}{E^2(b)E^2(c)} \left[1 + \frac{\text{var } a}{E^2(a)} + \frac{3 \text{ var } b}{E^2(b)} + \frac{3 \text{ var } c}{E^2(c)} + \frac{\text{var } d}{E^2(d)} \right. \\ \left. - \frac{4 \text{ cov}(a, b)}{E(a)E(b)} - \frac{4 \text{ cov}(a, c)}{E(a)E(c)} + \frac{4 \text{ cov}(a, d)}{E(a)E(d)} \right. \\ \left. + \frac{4 \text{ cov}(b, c)}{E(b)E(c)} - \frac{4 \text{ cov}(b, d)}{E(b)E(d)} - \frac{4 \text{ cov}(c, d)}{E(c)E(d)} \right] + \varepsilon$$

the expectations, variances and covariances to be used in these formulae correspond to the truncated distribution of W_I . As they depend of N_1, R_1, N_2, R_2 and N , they must be calculated for each particular experimental condition.

In Levene's condition (binomial marginal drawings, multinomial mating pair drawings), we obtain:

$$E(W_I) = 1 + \frac{1}{N} \left(\frac{1}{p'} + \frac{1}{p''} \right) + \varepsilon$$

$$\text{Var}(W_I) = \frac{1}{N} \left(\frac{1}{p} + \frac{1}{p'} + \frac{1}{p''} + \frac{1}{p'''} \right) - \frac{1}{N^2} \left(\frac{1}{p'} + \frac{1}{p''} \right)^2 + \varepsilon.$$

Although, this model is unrealistic, it is interesting to note that the expectation $E(W_I)$, calculated to a second order expansion, is frequency-dependent. The form of this frequency-dependence in the particular case of $P_1 = P_2 = P$ is shown in fig. 1. A simulation of this model, with elimination of indeterminate or infinite values of W_I , for $N = 10; 25; 50; 100; 500; 1000$ and $P_1 = P_2 = 0.1; 0.2; \dots; 0.9$, on the null hypothesis of no sexual selection and random mating, shows that this frequency-dependence exists and rapidly diminishes with increasing N . Practically, it disappears for $N > 500$, where Levene's formulae are valid.

FIG. 1.—The relation between $E(W_I)$ and the frequency (P) of A males (= frequency of A females), under the null hypothesis of no sexual selection and random matings (binomial marginal drawings and multinomial mating pair drawings). The number of mating pairs formed is indicated by N .

This frequency-dependence appears in the same way for the parameters of sexual selection:

$$\begin{aligned}
 E(W_M) &= 1 + \frac{1}{N} \left(\frac{1}{p''} + \frac{1}{p'''} \right) + \varepsilon \\
 \text{Var}(W_M) &= \frac{1}{N} \left(\frac{1}{p} + \frac{1}{p'} + \frac{1}{p''} + \frac{1}{p'''} \right) - \frac{1}{N^2} \left(\frac{1}{p''} + \frac{1}{p'''} \right)^2 + \varepsilon \\
 E(W_F) &= 1 + \frac{1}{N} \left(\frac{1}{p'} + \frac{1}{p'''} \right) + \varepsilon \\
 \text{Var}(W_F) &= \frac{1}{N} \left(\frac{1}{p} + \frac{1}{p'} + \frac{1}{p''} + \frac{1}{p'''} \right) - \frac{1}{N^2} \left(\frac{1}{p'} + \frac{1}{p'''} \right)^2 + \varepsilon.
 \end{aligned}$$

The same is true when the corresponding values are obtained from the hypergeometric model.

With:

$$\begin{aligned}
 U &= \frac{R_1 R_2}{N_1 N_2} - \frac{(R_1 - 1)(R_2 - 1)}{(N_1 - 1)(N_2 - 1)} & U' &= \frac{R_1(N_2 - R_2)}{N_1 N_2} - \frac{(R_1 - 1)(N_2 - R_2 - 1)}{(N_1 - 1)(N_2 - 1)} \\
 U'' &= \frac{(N_1 - R_1)R_2}{N_1 N_2} - \frac{(N_1 - R_1 - 1)(R_2 - 1)}{(N_1 - 1)(N_2 - 1)} \\
 U''' &= \frac{(N_1 - R_1)(N_2 - R_2)}{N_1 N_2} - \frac{(N_1 - R_1 - 1)(N_2 - R_2 - 1)}{(N_1 - 1)(N_2 - 1)},
 \end{aligned}$$

the Taylor expansion at the second order gives:

$$\begin{aligned}
 E(W_I) &= 1 + \frac{1}{N} \left(\frac{1}{p'} + \frac{1}{p''} \right) - \frac{(N-1)}{N} \left(\frac{U'}{p'} + \frac{U''}{p''} \right) + \varepsilon \\
 \text{Var}(W_I) &= \frac{1}{N} \left(\frac{1}{p} + \frac{1}{p'} + \frac{1}{p''} + \frac{1}{p'''} \right) - \frac{(N-1)}{N} \left(\frac{U}{p} + \frac{U'}{p'} + \frac{U''}{p''} + \frac{U'''}{p'''} \right) \\
 &\quad - \left[\frac{1}{N} \left(\frac{1}{p'} + \frac{1}{p''} \right) - \frac{(N-1)}{N} \left(\frac{U'}{p'} + \frac{U''}{p''} \right) \right]^2 + \varepsilon
 \end{aligned}$$

and analogous expressions for the expectations and variances of W_M and W_F . A frequency-dependence is then always observed, although the bias of $E(W_I)$ is less important, but the variance is enhanced.

Some investigators, using the observed variation of these parameters, have concluded that there is frequency-dependent selection. Before accepting such a conclusion, it is necessary to allow for the bias which we have already pointed out. This is the reason that makes us suspect that some claimed frequency-dependent phenomena could be artefacts introduced by this bias. This could be the case for the results of Muggleton (1979), Fontdevila and Mendez (1979) and others.

We see also that the goodness of a Taylor's approximation to the second order will depend on the proportions of the A and B individuals in each sex. According to the particular values of N , P_1 and P_2 , one could make use of Levene's formulae, our formulae, or the Taylor's expansions at even a higher order.

Sometimes, due to the limited number of pairs that can be observed together, and to the subsequent attainment of indeterminate or infinite

values of W_I , investigators add up the results of several experiments made under the same conditions. Putting $A = \sum_i ai$, $B = \sum_i bi$, ... and taking $W_I = AD/BC$, it is easy to show that, if the experiments are independent, the expressions given for $E(W_I)$ and $\text{Var}(W_I)$ are not modified. Such a procedure then diminishes the bias by enhancing N .

Hypothesis 4: The index W_I (resp. W_M and W_F) varies from 0 to $+\infty$. Its distribution is then very asymmetric. The use of a confidence interval based on the variance $\text{Var}(W_I)$ can then be criticized.

4. CONCLUSION

It appears that only a knowledge of the distribution of W_I (resp. W_M and W_F) could allow a correct definition of a confidence interval for this parameter. A sample of W_I values drawn from the whole set of W_I values, for a given N , has been obtained in the simulation we quoted above. The actual distributions show:

1—that W_I is not continuously distributed

2—that the distribution of W_I is not unimodal for small values of N .

It means that values of W_I with small probabilities can be inserted between values with high probabilities. It is only for large values of N (> 500) and not too low frequencies (P_1 or P_2) of each type A or B in males (or females), that a unimodal histogram can be obtained, the distribution remaining asymmetric. Thus, the use of a confidence interval for W_I obtained from a limited number of observed matings is misleading. At least, it is fair to ask if other quantitative indices measuring sexual isolation and selection would not be preferred to W and their related parameters. The literature on the interpretation of the 2×2 tables is considerable, but Edwards (1963) showed that a good measure of association must not be influenced by the variation of the numbers N_1 and N_2 of males and females, and as a consequence, has to take the form of a function of ad/bc . That is why the distance proposed by Schaffer (1968), explicitly a measure of "discrimination in mating", seems to be inadequate. Moreover, it is constrained by the same conditions as W relative to the numbers N , N_1 , N_2 , which must be large. Furthermore, its distribution is unknown.

Among the functions of W_I given by Kendall and Stuart (1973), are the Yule's coefficient of association, $(W_I - 1)/(W_I + 1)$, and the coefficient of colligation of the same author, $Y = (Z_I - 1)/(Z_I + 1)$. The index $L = \log W_I$ has also been advocated. Clearly, none of these coefficients can give more than W_I .

In conclusion, the peculiar properties of W indices make their use unsuitable for the measurement of sexual isolation or selection under the experimental conditions generally chosen by the investigators.

Acknowledgments.—We wish to thank Dr O'Donald for his helpful remarks, Professor C. Petit for providing us with a draft of H. Levene's calculations, and Professor C. Krimbas for useful comments during the writing of the manuscript.

5. REFERENCES

- EDWARDS, A. W. F. 1963. The measure of association in a 2×2 table. *J. Roy. Stat. Soc. London, B*, 126, 109-114.
- EHRMAN, L., AND PETIT, C. 1968. Genotype frequency and mating success in the Willistoni species group of *Drosophila*. *Evolution*, 22, 649-658.

- ELENS, A., VAN DEN HAUTE, J., AND DELCOUR, J. 1974. Sexual vigor and sexual selection in *Drosophila melanogaster*. *Evolution*, 27, 549-557.
- FONTDEVILA, A., AND MENDEZ, J. 1979. Frequency-dependent mating in a modified alloenzyme locus in *Drosophila pseudoobscura*. *Evolution*, 33, 634-640.
- HALDANE, J. B. S. 1956. The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Human Genet.*, 20, 309-311.
- KENDALL, M. G., AND STUART, E. 1973. *The advanced theory of statistics*. Griffin. London.
- MATHER, K. 1951. *The measurement of linkage in heredity*. Methuen and Co. London.
- MUGGLETON, J. 1979. Non-random mating in wild populations of polymorphic *Adalia bipunctata*. *Heredity*, 42, 57-65.
- O'DONALD, P. 1977. Theoretical aspects of sexual selection. *Theor. Pop. Biol.*, 12, 298-334.
- O'DONALD, P. 1978. A general model of mating behaviour with natural selection and female preference. *Heredity*, 40, 427-438.
- PETIT, C. 1951. Le rôle de l'isolement sexuel dans l'évolution des populations de *Drosophila melanogaster*. *Bull. Biol. Fr. Belg.*, 85, 392-418.
- PETIT, C., AND EHRLMAN, L. 1968. Le rôle de la sélection sexuelle dans l'évolution des populations: l'avantage des types rares dans le groupe willistoni (genre *Drosophila*, sous-genre *Sophophora*). *Bull. Biol.*, 102, 433-446.
- PETIT, C., KITAGAWA, O., AND TAKAMURA, T. 1976. Mating system between Japanese and French geographic strains of *Drosophila melanogaster*. *Jap. Journ. Genet.*, 51, 99-108.
- RIBO, G. 1977. Mécanisme de maintien de l'allèle *Bar* dans une population expérimentale de *Drosophila melanogaster*. *Arch. Zool. Exp.*, 118, 375-389.
- SCHAFFER, H. E. 1968. A measure of discrimination in mating. *Evolution*, 22, 125-129.
- SPIESS, E. B., AND SCHWER, W. A. 1978. Minority mating advantage of certain eye color mutants of *Drosophila melanogaster*. I. Multiple-choice and single-female tests. *Behavior Genetics*, 8, 155-168.
- WHITE, H. C., AND GRANT, B. 1977. Olfactory cues as a factor in frequency-dependent mate selection in *Mormoniella vitripennis*. *Evolution*, 31, 829-835.