## II. THE USE OF THE PRODUCT FORMULA FOR THE ESTIMATION OF LINKAGE IN INTERCROSSES WHEN DIFFERENTIAL VIABILITY IS PRESENT

NORMAN T. J. BAILEY

The offspring of intercrosses between double heterozygotes of the constitution AB/*ab* (coupling) or A*b*/*a*B (repulsion) fall into four phenotypic classes: AB, A*b*, *a*B and *ab*. If viability is unaffected the expected numbers in these four classes are, respectively, $2+\theta$, $1-\theta$, $1-\theta$, and $\theta$; where, if $p$ and $p'$ are the recombination fractions for male and female gametogenesis, $\theta = pp'$ for matings in repulsion and $\theta = (1-p)(1-p')$ for matings in coupling. Suppose the observed and expected numbers are as follows:—

TABLE 1

| | AB | A*b* | *a*B | *ab* | |
|---|---|---|---|---|---|
| Expected . . . | $2+\theta$ | $1-\theta$ | $1-\theta$ | $\theta$ | $\times\frac{1}{4}n$ |
| Observed . . . | $a$ | $b$ | $c$ | $d$ | |

where
$$a+b+c+d = n \qquad (1)$$

Now as Fisher and Balmakund (1928) have shown, the consistent estimate T of the parameter $\theta$, given by

$$\frac{ad}{bc} = \frac{\mathrm{T}(2+\mathrm{T})}{(1-\mathrm{T})^2} \qquad (2)$$

is an efficient estimate. For we have the maximum likelihood equation

$$\frac{\partial \mathrm{L}}{\partial \theta} \equiv \frac{a}{2+\theta} - \frac{b+c}{1-\theta} + \frac{d}{\theta} = 0 \qquad (3)$$

$$\therefore\ n\hat{\theta}^2 - (a-2b-2c-d)\hat{\theta} - 2d = 0 \qquad (4)$$

Further,
$$\operatorname{var} \hat{\theta} = -1/\mathrm{E}\frac{\partial^2 \mathrm{L}}{\partial \theta^2} = \frac{2\theta(1-\theta)(2+\theta)}{n(1+2\theta)} \qquad (5)$$

If we now use the formula:

$$\operatorname{var} \mathrm{T} = \Sigma m\left(\frac{\partial \mathrm{T}}{\partial a}\right)^2 - n\left(\frac{\partial \mathrm{T}}{\partial n}\right)^2 \qquad (6)$$

where the $m$'s are the expectations, to calculate the variance of T given by (2), it will be found equal to var $\hat{\theta}$ in (5). T is thus efficient.

### Differential viability in one pair of factors

Let us now suppose that the viability of A phenotypes relative to *a* phenotypes is $u$. This problem has been treated by Fisher (1939), where he shows that the product formula can still be used. He gives an expression for the variance of the corresponding estimate of $\theta$ involving $u$, which is in turn easily estimated from the data. It will be seen below that this estimate of $\theta$ does not utilise quite all the information available but for moderate disturbances of viability very little is lost.

The expected and observed numbers are given by :—

TABLE 2

| | AB | A$b$ | $a$B | $ab$ | |
|---|---|---|---|---|---|
| Expected . . | $u(2+\theta)$ | $u(1-\theta)$ | $(1-\theta)$ | $\theta$ | $\times n/(3u+1)$ |
| Observed . . | $a$ | $b$ | $c$ | $d$ | |

Now
$$e^{L} \propto \frac{u^{a+b}(2+\theta)^{a}(1-\theta)^{b+c}\theta^{d}}{(3u+1)^{n}} \quad . \quad . \quad . \quad (7)$$

$$\therefore \frac{\partial L}{\partial \theta} \equiv \frac{a}{2+\theta} - \frac{b+c}{1-\theta} + \frac{d}{\theta} = 0 \quad . \quad . \quad (3 \text{ bis})$$

which is identical with (3)

and
$$\frac{\partial L}{\partial u} \equiv \frac{a+b}{u} - \frac{3n}{3u+1} = 0 \quad . \quad . \quad . \quad (8)$$

From (8) we obtain :
$$\frac{3u}{a+b} = \frac{3u+1}{n} = \frac{1}{c+d} \quad . \quad . \quad . \quad . \quad (9)$$

We also have from (3 bis) and (8) :

$$I_{\theta\theta} = -E\frac{\partial^2 L}{\partial \theta^2} = \frac{n}{3u+1}\left\{\frac{u}{2+\theta} + \frac{1+u}{1-\theta} + \frac{1}{\theta}\right\} = \frac{n[2+\theta(3u+1)]}{(3u+1)\theta(1-\theta)(2+\theta)} \quad (10)$$

$$I_{\theta u} = -E\frac{\partial^2 L}{\partial \theta \delta u} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$I_{uu} = -E\frac{\partial^2 L}{\partial u^2} = \frac{3n}{u(3u+1)^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

(11) shows that the estimates $\hat{u}$ and $\hat{\theta}$ are uncorrelated ; their variances are therefore $I_{uu}^{-1}$ and $I_{\theta\theta}^{-1}$. By using the transformation

$$u = \frac{1}{3}\tan^2\alpha \quad . \quad . \quad . \quad . \quad (13)$$

we can change to a variable $\alpha$ whose sampling variance $1/4n$ depends only on the sample number.

Now consider the consistent product formula given by (2), *i.e.*

$$\frac{ad}{bc} = \frac{T(2+T)}{(1-T)^2}$$

The easiest way to calculate var T is to write $F = \log \frac{ad}{bc}$ and then to use (6) to evaluate var F. For large samples we then have :

$$\text{var } T = \text{var } F \Big/ \left(\frac{\partial F}{\partial T}\right)^2_{T=\theta} = \frac{3u+1}{4nu} \cdot \frac{\theta(1-\theta)(2+\theta)}{(1+2\theta)^2}[2u+\theta(3+u)] \quad . \quad (14)$$

Referring to (10) we see that the efficiency E is given by

$$E = \frac{4u(1+2\theta)^2}{[2+\theta(3u+1)][2u+\theta(3+u)]} \quad . \quad . \quad . \quad (15)$$

It will be noticed that substituting $1/u$ for $u$ leaves E unchanged. If we write $x$ and $y$ for the numerator and denominator in (15) :

$$y-x = 3\theta(2+\theta)(1-u)^2 \quad . \quad . \quad . \quad (16)$$

Thus E is always less than unity except when $\theta = 0$, or $u = 1$. Differentiating (15) with respect to $\theta$ gives :

$$\frac{\partial E}{\partial \theta} = \frac{24u(u-1)^2(1+2\theta)(\theta-1)}{[2+\theta(3u+1)]^2[2u+\theta(3+u)]^2} \quad . \quad . \quad (17)$$

So that for a given value of $u$, not equal to unity, E is always decreasing in the range $0 \leqslant \theta < 1$.

Similarly :

$$\frac{\partial E}{\partial u} = \frac{12(1+2\theta)^2\theta(2+\theta)(1-u^2)}{[2+\theta(3u+1)]^2[2u+\theta(3+u)]^2} \quad . \quad . \quad (18)$$

For a given value of $\theta$, not equal to zero, E is always decreasing in the range $1 < u \leqslant \infty$, or $1 > u \geqslant 0$. A few values of E for different $\theta$ and $u$ are given in the following table :—

TABLE 3

| | $u$: ... / 1·00 | 1·50 / 0·67 | 2·00 / 0·50 | 2·50 / 0·40 | 3·00 / 0·33 | $p$ (or $\sqrt{pp'}$) | |
|---|---|---|---|---|---|---|---|
| $\theta$ 1·00 | 1·00 | 0·96 | 0·89 | 0·82 | 0·75 | 0·0 | Coupling |
| $\theta$ 0·25 | 1·00 | 0·97 | 0·91 | 0·86 | 0·80 | 0·5 | |
| $\theta$ 0·00 | 1·00 | 1·00 | 1·00 | 1·00 | 1·00 | 0·0 | Repulsion |

For matings in repulsion $\theta$ increases from 0 to $\frac{1}{4}$ as the recombination fraction increases from 0 to $\frac{1}{2}$. For matings in coupling the corresponding range of $\theta$ is from 1 to $\frac{1}{4}$. Given $u$, the loss of efficiency is more serious in the coupling phase than in the repulsion phase, being worst for close linkage in coupling. But for values of $u$ lying between $\frac{1}{2}$ and 2 the loss is never more than 11 per cent. ; and for values between $\frac{2}{3}$ and $1\frac{1}{2}$ does not exceed 4 per cent.

### Differential viability in both pairs of factors

Now suppose that, in addition to the differential viability of A with respect to $a$, there is also a differential viability, $v$, of B phenotypes relative to $b$. Assume the two effects are independent. The observed and expected numbers are :—

TABLE 4

| | AB | A$b$ | $a$B | $ab$ | |
|---|---|---|---|---|---|
| Expected . . | $uv(2+\theta)$ | $u(1-\theta)$ | $v(1-\theta)$ | $\theta$ | $\times n/\{uv(2+\theta)+(u+v)(1-\theta)+\theta\}$ |
| Observed . . | $a$ | $b$ | $c$ | $d$ | |

$$\therefore e^{L} \propto \frac{u^{a+b}v^{a+c}\theta^{d}(1-\theta)^{b+c}(2+\theta)^{a}}{[uv(2+\theta)+(u+v)(1-\theta)+\theta]^{n}} \quad . \quad . \quad (19)$$

It will simplify the treatment if we write $m_1$, $m_2$, $m_3$ and $m_4$ for the four expected numbers ; also let us write

$$D = uv(2+\theta)+(u+v)(1-\theta)+\theta \quad . \quad . \quad . \quad (20)$$

Differentiating (19) logarithmically gives :

$$\frac{\partial L}{\partial \theta} = \frac{a-m_1}{2+\theta} - \frac{(b+c)-(m_2+m_3)}{1-\theta} + \frac{d-m_4}{\theta} \quad . \quad . \quad (21)$$

$$\frac{\partial L}{\partial u} = \frac{(a+b)-(m_1+m_2)}{u} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$$\frac{\partial L}{\partial v} = \frac{(a+c)-(m_1+m_3)}{v} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The maximum likelihood equations are given by equating each of (21), (22) and (23) to zero. After a little manipulation we obtain :

$$\left.\begin{aligned} a = m_1 = \frac{nuv(2+\theta)}{D} ;\ b = m_2 = \frac{nu(1-\theta)}{D} \\ c = m_3 = \frac{nv(1-\theta)}{D} ;\ d = m_4 = \frac{n\theta}{D} \end{aligned}\right\} \quad . \quad . \quad (24)$$

From (24) we can recover the familiar product formula :

$$\frac{ad}{bc} = \frac{\theta(2+\theta)}{(1-\theta)^2} \quad . \quad . \quad . \quad . \quad (25)$$

which now yields the maximum likelihood estimate of $\theta$. We can also obtain from (24) quadratics for $u$ and $v$ :

$$3cdu^2+2bcu-ab = 0 \quad . \quad . \quad . \quad . \quad (26)$$

and

$$3dbv^2+2bcv-ac = 0 \quad . \quad . \quad . \quad . \quad (27)$$

$$\therefore \hat{u} = \frac{b}{3d}\left[\left(1+3\frac{ad}{bc}\right)^{\frac{1}{2}} - 1\right] \quad . \quad . \quad . \quad (28)$$

$$\hat{v} = \frac{c}{3d}\left[\left(1+3\frac{ad}{bc}\right)^{\frac{1}{2}} - 1\right] \quad . \quad . \quad . \quad (29)$$

Note that

$$\hat{v} = \frac{c}{b}\hat{u} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

We must now derive the sampling variance of $\hat{\theta}$. As before we take $F = \log \frac{ad}{bc}$ and employ (6) to calculate the sampling variance of F. Then :

$$\text{var } \hat{\theta} = \text{var F} \Big/ \left(\frac{\partial F}{\partial \theta}\right)^2$$

$$= \frac{\theta(1-\theta)(2+\theta)}{4nuv(1+2\theta)^2} [uv(2+\theta)+(u+v)(1-\theta)+\theta]$$
$$\times [\theta(1-\theta)+(u+v)\theta(2+\theta)+uv(2+\theta)(1-\theta)] \quad . (31)$$

With the help of (24) we can write (31) a little more conveniently in the form :

$$\text{var } \hat{\theta} = \frac{\theta^2(2+\theta)^2}{4ad(1+2\theta)^2} [(a+d)+2\theta\{(b+c)-(a+d)\}+n\theta^2] \quad . (32)$$

I am grateful to Professor R. A. Fisher for pointing out that if we wish to judge the significance of departures of $v$, say, from unity, admitting linkage and one disturbed viability, then we can use the ordinary expression for the $\chi^2$ which measures the agreement between the observed and the expected numbers. Of the three available degrees of freedom two are used up in the estimation of $\theta$ and $u$. Referring to table 2 for the expectations, and using (9) we have :—

TABLE 5

| | AB | A$b$ | $a$B | $ab$ |
|---|---|---|---|---|
| Expected . . . | $\frac{1}{3}(a+b)(2+\theta)$ | $\frac{1}{3}(a+b)(1-\theta)$ | $(c+d)(1-\theta)$ | $(c+d)\theta$ |
| Observed . . . | $a$ | $b$ | $c$ | $d$ |

where $\theta$ is given by (3 bis) or (4).

$$\therefore \chi^2 = \frac{3a^2}{(a+b)(2+\theta)} + \frac{3b^2}{(a+b)(1-\theta)} + \frac{c^2}{(c+d)(1-\theta)} + \frac{d^2}{(c+d)\theta} - n \quad . (33)$$

The following alternative forms are also due to Professor Fisher : It follows from (3 bis) that :

$$\theta\{a(1-\theta)-b(2+\theta)\} = (2+\theta)\{c\theta-d(1-\theta)\} = k, \text{ say.} \quad . (34)$$

and
$$\frac{a(1-\theta)-b(2+\theta)}{2+\theta} = \frac{c\theta-d(1-\theta)}{\theta} = \partial, \text{ say.} \quad . \quad . (35)$$

Hence (33) can also be written

$$\chi^2 = \frac{k^2}{\theta(1-\theta)(2+\theta)} \left\{\frac{1}{(a+b)\theta} + \frac{1}{(c+d)(2+\theta)}\right\} = \frac{\partial^2}{1-\theta} \left\{\frac{2+\theta}{a+b} + \frac{\theta}{c+d}\right\} \quad (36)$$

## Summary

In intercrosses the product formula yields an estimate of $\theta$, which with undisturbed viability is fully efficient, and, with a single disturbed viability, results in the loss of a small amount of information for moderate disturbances. With two disturbed viabilities, however, the estimate is again fully efficient and is in fact then identical with the maximum likelihood solution. It may also be noted that the maximum likelihood solution for $\theta$ is the same in both the undisturbed case and the case of a single disturbed viability.

## References

FISHER, R. A., AND BALMAKUND, B. 1928.
The estimation of linkage from the offspring of selfed heterozygotes.
*J. Genet.*, *20*, 79-92.

FISHER, R. A. 1939.
The precision of the product formula in the estimation of linkage.
*Ann. Eug.*, *9*, 50-54.

# III. A METHOD OF ALLOWING FOR DIFFERENTIAL VIABILITY IN ESTIMATING LINKAGE FROM BACKCROSS MATINGS IN COUPLING ONLY OR REPULSION ONLY

NORMAN T. J. BAILEY

A coupling mating of the type : AB/*ab* = *ab*/*ab*, gives rise to four kinds of offspring which are, phenotypically : AB, *ab* (parentals) and A*b*, *a*B (recombinants). It is usual to group the two parental types and the two recombinant types to give a pair of observed numbers. Corresponding results may be obtained for matings in repulsion. It has been shown by Fisher (1935-47) that the joint use of coupling and repulsion data can be made to yield maximum likelihood estimates of (*a*) the linkage between the two loci and (*b*) the relative viability of the two groups AB, *ab* and A*b*, *a*B.

Often, however, results are available for matings in coupling only or repulsion only. We can still take into account the effects of differential viability by considering the numbers observed in all four classes : AB, *ab*, A*b* and *a*B, on the assumption that the presence of each dominant gene modifies the expected values by a certain factor and that the two viability effects are independent.

Let the recombination fraction be $p$ ; the factors corresponding to the presence of A and B be $u$ and $v$ respectively ; and the sample number be $n$. Let the observed and expected numbers for matings in coupling be given by the following table :—

| | AB | *ab* | A*b* | *a*B | |
|---|---|---|---|---|---|
| Expected . . | $nuvq$ | $nq$ | $nup$ | $nvp$ | $\div(uvq+q+up+vp)$ |
| Observed . . | $a$ | $b$ | $c$ | $d$ | |

where $$p+q = 1 \quad . \quad . \quad . \quad . \quad . \quad (1)$$
and $$a+b+c+d = n \quad . \quad . \quad . \quad . \quad . \quad (2)$$

For matings in repulsion we merely interchange $p$ and $q$.