17 March 1931.

J.B.S. Haldane, Esq., M.A.,
John Innes Horticultural Inst.,
Hoblyn Road,
Hertford Park, N.W. 19.

Dear Haldane,

Your problem is a very good example of the general use of maximum likelihood as explained in Section 57.1 of my book.

If you are satisfied in respect of visibility etc., that the data ought to be combined, you proceed to write down the expectation and observed frequencies in each set of experiments:

(a) Backcrossed females

Expectation \( \frac{1}{2} n \left\{ 1 - \rho, \quad \rho, \quad \rho, \quad 1 - \rho \right\} \)

Observed \( a, \quad b, \quad c, \quad d \)

(b) Backcrossed males

Expectation \( \frac{1}{2} n \left\{ 1 - q, \quad q, \quad q, \quad 1 - q \right\} \)

Observed \( a_2, \quad b_2, \quad c_2, \quad d_2 \)

(Y) \( F_2 \) Coupling

Expectation \( \frac{1}{4} n_3 \left\{ 2 + (1 - \rho)(1 - q), \quad 1 - (1 - \rho)(1 - q), \quad 1 - (1 - \rho)(1 - q), \quad (1 - \rho)(1 - q) \right\} \)

Observed \( a_3, \quad b_3, \quad c_3, \quad d_3 \)

(8) \( F_2 \) Repulsion \( \frac{1}{4} n_4 \left\{ 2 + \rho q, \quad 1 - \rho q, \quad 1 - \rho q, \quad \rho q \right\} \)

Observed \( a_4, \quad b_4, \quad c_4, \quad d_4 \)
Thence the logarithm of the likelihood is

\[(a_x + d_x) \log (1 - \beta) + (b_x + c_x) \log \beta + (a_y + d_y) \log (1 - q) + (b_y + c_y) \log q + a_z \log \left\{ 1 + (1 - \beta)(1 - q) \right\} + (b_z + c_z) \log \left\{ 1 - (1 - \beta)(1 - q) \right\} + a_z \log (1 - \beta) + (b_z + c_z) \log q \]

and the two equations for \( \beta \) and \( q \) are found by differentiation,

\[\frac{\partial L}{\partial \beta} = \frac{a_x + d_x + d_y}{1 - \beta} + \frac{b_x + c_x + d_y}{\beta} - \frac{a_y (1 - q)}{2 + (1 - \beta)(1 - q)} - \frac{(b_y + c_y)(1 - q)}{1 - (1 - \beta)(1 - q)} + \frac{a_w q}{2 + \beta q} - \frac{(b_w + c_w) q}{1 - \beta q} = 0\]

and

\[\frac{\partial L}{\partial q} = \frac{a_x + d_x + d_y}{1 - q} + \frac{b_x + c_x + d_y}{q} - \frac{a_y (1 - \beta)}{2 + (1 - \beta)(1 - q)} - \frac{(b_y + c_y)(1 - \beta)}{1 - (1 - \beta)(1 - q)} + \frac{a_w \beta}{2 + \beta q} - \frac{(b_w + c_w) \beta}{1 - \beta q} = 0\]

will in which you notice that all weighting is taken care of automatically by means of the observed frequencies.

Solution proceeds numerically, using probably the values from (x) and (y) as first approximations.

To systematise the process, and also to estimate the precision of the solution, it will be useful to calculate

\[-\frac{\partial^2 L}{\partial \beta^2}, \quad -\frac{\partial^2 L}{\partial \beta \partial q}, \quad -\frac{\partial^2 L}{\partial q^2}\]

for the trial values of \( \beta \) and \( q \). There is no need to trouble about algebraic simplification.

If \( \beta_o, q_o \) are the trial values, then the corrections \( \delta \beta_o, \delta q_o \) are given by

\[\delta \beta_o \left( -\frac{\partial^2 L}{\partial \beta_o^2} \right) + \delta q_o \left( -\frac{\partial^2 L}{\partial \beta_o \partial q_o} \right) = \frac{\partial L}{\partial \beta_o}\]
\[ \delta \rho \left( -\frac{\partial^2 L}{\partial \rho \partial q} \right) + \delta q \left( -\frac{\partial^2 L}{\partial \rho \partial q^2} \right) = \frac{\partial L}{\partial q} \]

giving a new solution

\[ \rho = \rho_0 + \delta \rho, \quad q = q_0 + \delta q \]

which will I think be efficient, though capable of further improvement in the same way. If \( a, b, c \) stand for the second differential coefficients of \( L \), with changed sign, then \( \frac{1}{a} \) gives the sampling variance of \( \rho \) for given observed \( q \), while the total sampling variance of \( \rho \) is \[ \frac{c}{ac - b^2} \], and the sampling covariance of \( \rho \) and \( q \) is \[ b/(ac - b^2) \]. Consequently, the sampling variance of \( \rho - q \) is

\[ \frac{a + c - 2b}{ac - b^2} \]

and you will probably want this. A point which might be worth while following through in the algebra is the amount of information respecting \( \rho - q \), i.e. \( (ac - b^2)/(a + c - 2b) \) to be expected from samples of \( n_1, n_2, n_3, \) and \( n_4 \), for this would show how much each type of experiment contributed to the total amount of information available.

Do you believe the Berlin tale? The fallacies of Edin's work on Stockholm are fairly easy to see, but I have not looked at the Berlin stuff.

Yours sincerely,

R. A. Kim