THE FIDUCIAL LIMITS OF TETRAD-RANK FREQUENCIES*

DONALD R. CHARLES[1]

Department of Biology, University of Rochester

Received March 11, 1957

W EINSTEIN'S equations (1936) yield *maximum likelihood* estimates of the regional frequency distribution of crossovers in the tetrads from which a given sample of strands was derived, one per tetrad. But clearly, as with all estimates of universe parameters from sample statistics, *other* values within a certain range cannot be rejected; that range can be ascertained, for each tetrad-rank separately, from the standard error computed by a method also given by WEINSTEIN.

The standard errors of the *separate* frequencies do not, however, define the range of non-rejectable values of the rank frequencies *jointly*, where there are more than two ranks. It is with the calculation of the joint range that the present note is concerned.

Only one special case will be considered here: the frequencies of tetrads of rank $0, 1, 2, \cdots, k$ (or, further on, $k + 1$) with regard not to the regional location of the crossovers but only to the total number per tetrad, estimated from $(k + 1)$ observed strand-rank frequencies, assuming random recurrence and no sister-strand crossing over.

All possible sets of $(k + 1)$ tetrad frequencies, expressed in decimal fraction, may be thought of as lying within a k-dimensional, unit-edge hypercube, occupying a "corner" of volume $1/k!$. Somewhere within the corner is the point representing Weinstein's maximum likelihood estimate of the universe tetrad-rank frequencies for a given sample.

Now consider any other point within the corner, any other set of universe rank frequencies. Around it lie successive hyperellipsoidal shells (with increasingly flattened areas near the surfaces of the corner), each containing a finite set of points representing equally-likely sets of rank frequencies in samples from the given universe. The likelihood decreases outward from shell to shell. The sum of the likelihoods outside a given shell falls from 1 at the universe point to slightly above 0 at the plane surfaces of the corner (possibly somewhat higher at one of the surfaces).

Suppose next that we locate all points within the corner whose equal-likelihood sample shells passing through the maximum likelihood point have some particular outside likelihood sum, P. These points will themselves form a hyperellipsoidal shell, with possibly one or more flattened regions. Our problem is to locate that shell, for P generally somewhere from .01 to .05.

But the complete specifications of such a fiducial limit shell would not be very useful, because incomprehensible in any down-to-earth sense. Obviously then we have to confine our attention to particular points within the shell, to specific sets of

possible universe frequencies: probably, by analogy with the fiducial limits of a single parameter, to some *pair* of points in opposite directions from the maximum likelihood point, crudely like the two apices of a football.

Several such pairs could be defined, at least where four or more strand ranks are observed. One pair that would seem to be of special interest is that point within the fiducial-limit shell at which the *average tetrad-rank* is *highest*, and that point at which it is *lowest*.

The conditions we have imposed are, altogether, as follows:

$$1. \ \sum_{i=0}^{k} T_i = 1 \qquad 2. \ \sum_{i=0}^{k} (s_i - \sigma_i)^2/\sigma_i = G/N \qquad 3. \ \frac{\partial}{\partial T_i} \, i_T = 0$$

Here T_i represents a universe frequency of tetrads of rank i; σ_i, a universe strand frequency; s_i, an observed strand frequency; N, total strands in a sample; k, the highest observed strand rank; G, the value of χ^2 for k degrees of freedom and some specified fiducial probability, ordinarily .01 to .05.

The relation of the strand frequencies (σ's) to the tetrad frequencies (T's), assuming that there is no crossing over between sister strands but that otherwise association of strands is random, has been given by WEINSTEIN (1936) as follows:

$$\sigma_0 = T_0 + \tfrac{1}{2} T_1 + \tfrac{1}{4} T_2 + \tfrac{1}{8} T_3 + \tfrac{1}{16} T_4 \cdots$$

$$\sigma_1 = \qquad \tfrac{1}{2} T_1 + \tfrac{2}{4} T_2 + \tfrac{3}{8} T_3 + \tfrac{4}{16} T_4 \cdots$$

$$\sigma_2 = \qquad\qquad\quad \tfrac{1}{4} T_2 + \tfrac{3}{8} T_3 + \tfrac{6}{16} T_4 \cdots$$

$$\sigma_3 = \qquad\qquad\qquad\qquad \tfrac{1}{8} T_3 + \tfrac{4}{16} T_4 \cdots$$

$$\cdot\cdot \qquad\qquad\qquad\qquad \cdot\cdot \qquad \cdot\cdot \quad \cdots$$

$$\cdot\cdot \qquad\qquad\qquad\qquad \cdot\cdot \qquad \cdot\cdot \quad \cdots$$

$$\overline{\Sigma\sigma_i = T_0 + \quad T_1 + \quad T_2 + \quad T_3 + \quad T_4 \cdots = \Sigma T_i = 1}$$

$$i_\sigma = \Sigma i\sigma_i = \tfrac{1}{2} T_1 + \tfrac{2}{2} T_2 + \tfrac{3}{2} T_3 + \tfrac{4}{2} T_4 \cdots = (\tfrac{1}{2})\Sigma i T_i = \tfrac{1}{2} i_T$$

Since condition 2 has to be given in terms of s_i, and hence σ_i, the other two conditions can be imposed more conveniently by bringing them into the same terms:

$$1'. \ \sum_{i=0}^{k} \sigma_i = 1 \qquad 3'. \ \frac{\partial}{\partial T} \, i_\sigma = 0$$

Condition 3' is equivalent to 3 because $i_\sigma = i_T/2$ for all sets of tetrad-rank frequencies and the strand frequencies calculated from them.

We now express σ_0 and σ_1 in terms of i_σ and σ_2 to σ_k (from condition 1' and the definition of i_σ). We then derive i_σ as a function of σ_2 to σ_k by means of condition 2.

$$\sigma_0 = 1 - i_\sigma + \sum_{i=2}^{k} (i - 1)\sigma_i \, ; \qquad \sigma_1 = i_\sigma - \sum_{i=2}^{k} i\sigma_i$$

$$\frac{G}{N} + 1 - \sum_{i=2}^{k} \frac{s_i^2}{\sigma_i} = \frac{s_0^2}{1 + \sum_{i=2}^{k} (i - 1)\sigma_i - i_\sigma} - \frac{s_1^2}{\sum_{i=2}^{k} i\sigma_i - i_\sigma}$$

Abbreviating the latter equation we have the following, in which the meanings of A, B and C will be obvious by comparison.

$$A = \frac{s_0{}^2}{B - i_\sigma} - \frac{s_1{}^2}{C - i_\sigma}$$

$$2i_\sigma = \left[-\frac{s_0{}^2 - s_1{}^2}{A} + B + C \right] \pm \left[\left(\frac{s_0{}^2 + s_1{}^2}{A} - B + C \right)^2 - \frac{4s_0{}^2 s_1{}^2}{A^2} \right]^{1/2}$$

Both roots satisfy the equation but if the values $A = \frac{s_0{}^2}{\sigma_0} + \frac{s_1{}^2}{\sigma_1}$, $B = i_\sigma + \sigma_0$, $C = i_\sigma - \sigma_1$ are substituted in the formula for i_σ, it will be found that it becomes an identity on using the plus sign when $s_0{}^2 \sigma_1{}^2 - s_1{}^2 \sigma_0{}^2 \geqq 0$, and the minus sign when $s_0{}^2\sigma_1{}^2 - s_1{}^2\sigma_0{}^2 < 0$.

$$2\frac{\partial i_\sigma}{\partial T_i} = \left[\frac{s_0{}^2 - s_1{}^2}{A^2} \frac{\partial A}{\partial T_i} + \frac{\partial B}{\partial T_i} + \frac{\partial C}{\partial T_i} \right] - \left[\left(\frac{s_0{}^2 + s_1{}^2}{A} - B + C \right)^2 - \frac{4s_0{}^2 s_1{}^2}{A^2} \right]^{-1/2}$$

$$\cdot \left\{ \frac{1}{A^2} \left[\frac{(s_0{}^2 - s_1{}^2)^2}{A} - (B - C)(s_0{}^2 + s_1{}^2) \right] \frac{\partial A}{\partial T_i} \right.$$

$$\left. + \left[\frac{s_0{}^2 + s_1{}^2}{A} - B + C \right] \left[\frac{\partial B}{\partial T_i} - \frac{\partial C}{\partial T_i} \right] \right\}$$

Setting this expression equal to 0, for i_σ maximal or minimal, and resubstituting for A, B and C in terms of σ_0, σ_1 and i_σ gives, after some simplification:

$$\frac{\partial A}{\partial T_i} + \left(\frac{s_0}{\sigma_0} \right)^2 \frac{\partial B}{\partial T_i} - \frac{s_1{}^2}{\sigma_1} \frac{\partial C}{\partial T_i} = 0 \qquad \cdots (1)$$

The partial derivatives of A, B and C with respect to T_i can be obtained on substituting their values in terms of σ_2 to σ_k. Thus

$$\frac{\partial A}{\partial T_i} = \sum_{j=2}^{k} \left(\frac{s_j}{\sigma_j} \right)^2 \frac{\partial \sigma_j}{\partial T_i}, \qquad \frac{\partial B}{\partial T_i} = \sum_{j=2}^{k} (j - 1) \frac{\partial \sigma_j}{\partial T_i} \qquad \text{and} \qquad \frac{\partial C}{\partial T_i} = \sum_{j=2}^{k} j \frac{\partial \sigma_j}{\partial T_i}.$$

It may be seen from Weinstein's equations that elimination of σ_0 and σ_1 automatically eliminates T_0 and T_1 and that $\frac{\partial \sigma_2}{\partial T_2} = \frac{1}{4}$, $\frac{\partial \sigma_2}{\partial T_3} = \frac{3}{8}$, $\frac{\partial \sigma_3}{\partial T_3} = \frac{1}{8}$, etc.

From the partial derivative of equation (1) with respect to T_2 we obtain

$$\left(\frac{s_2}{\sigma_2} \right)^2 = 2 \left(\frac{s_1}{\sigma_1} \right)^2 - \left(\frac{s_0}{\sigma_0} \right)^2 \qquad \cdots (2)$$

Taking the partials with respect to T_3, and subtracting (2) multiplied by $\partial \sigma_2 / \partial T_3$, gives

$$\left(\frac{s_3}{\sigma_3} \right)^2 = 3 \left(\frac{s_1}{\sigma_1} \right)^2 - 2 \left(\frac{s_0}{\sigma_0} \right)^2 \qquad \cdots (3)$$

So we may proceed, through T_k.

Now, substituting the values of $\sigma_2 \cdots \sigma_k$ from (2), (3), etc. into conditions 1' and 2, we obtain

$$\left(\frac{\sigma_0}{s_0}\right) \sum_{i=0}^{k} s_i (ix^2 - i + 1)^{-1/2} = 1 \qquad \cdots (4)$$

$$\left(\frac{s_0}{\sigma_0}\right) \sum_{i=0}^{k} s_i (ix^2 - i + 1)^{1/2} = \frac{G}{N} + 1 \qquad \cdots (5)$$

where x represents $s_1 \sigma_0 / s_0 \sigma_1$.

Eliminating σ_0 between (4) and (5) gives the following, with x as the only unknown. There are two roots, in the neighborhood of 1,

$$\sum_{i=0}^{k} s_i (ix^2 - i + 1)^{-1/2} \cdot \sum_{i=0}^{k} s_i (ix^2 - i + 1)^{1/2} = \frac{G}{N} + 1 \qquad \cdots (6)$$

approximately $[1 \pm (g/NV)^{1/2}]$ where V is the variance of the *observed* strand ranks. The larger root relates to the lower fiducial limit; the smaller, to the upper. Each can be computed, to any desired degree of accuracy, by some successive approximation method like the Newton-Raphson (WHITTAKER and ROBINSON 1944). The result can be substituted into either (4) or (5) to find σ_0 which, in turn, with x gives σ_1. Then $\sigma_2 \cdots \sigma_k$ can be calculated from (2), (3), etc. Finally the two sets of values of $\sigma_0 \cdots \sigma_k$ are put into Weinstein's equations to determine the lower and upper fiducial limits of the tetrad frequencies, as defined above.

The whole process is rather tedious, especially evaluating x. A close enough approximation, for most purposes, can be gotten from the first five terms of a Taylor-expansion of (6). Using δ to represent $(x - 1)$, T for the ratio of third central moment to variance of observed strand ranks, and F for the ratio of fourth central moment to variance,

$$\delta^2 \{1 - [2T + 4\bar{\imath}_s - 1]\delta$$
$$+ [3.75F + 3(4\bar{\imath}_s - 1)T + 6\bar{\imath}_s(2\bar{\imath}_s - 1) + 0.25(1 - 3V)]\delta^2\} = G/NV \qquad \cdots (7)$$

This too has to be handled by successive approximation but seems generally to require much less effort than (6).

An example of the outcome of such fiducial limit calculations is shown in table 1. It will be noticed that upper and lower limits are not equally above and below the observed. The asymmetry is not very large, in the present case, because N is very large; it may, however, be considerable in smaller experiments.

The upper fiducial limits we have obtained are not wholly satisfactory: we have rejected universes with some tetrads of rank higher than any in the observed strands. Judging from the sharp decline in frequency with rank, in the higher strand ranks, we need to worry only about tetrads of one higher rank than the highest observed strands. For these allowance can be readily made, as follows: We calculate, for the sample with which we are concerned, the value of σ_{k+1} for which the likelihood of $s_{k+1} = 0$ is some assigned value, generally the same as the fiducial probability corresponding to our chosen value of G. (This will usually violate the rule that, in a χ^2 calculation, no expected class frequency should be less than 5. If it is preferred to adhere to that rule, σ_{k+1} may be taken as $5/N$ which is equivalent to choosing a likelihood of .0067 for $s_{k+1} = 0$.) The effects of allowing $\sigma_{k+1} \neq 0$ are: i. to replace G by

TABLE 1

Fiducial limits of strand and tetrad rank-frequencies

Rank			0	1	2	3
Strand frequency						
observed			6607	7555	1913	61
lower limit, P =	.01, equation	7	6795.6	7463.7	1820.7	56.1
lower limit,	.05,	7	6762.5	7480.5	1836.2	56.9
lower limit,	.05,	6	6762.3	7480.6	1836.2	56.9
upper limit,	.05,	7	6456.1	7619.2	1995.4	65.4
upper limit,	.01,	7	6425.4	7630.7	2012.9	67.1
Tetrad frequency						
lower limit,	.01,	7	1096.6	7981.1	6609.8	448.6
lower limit,	.05,	7	1061.3	7957.6	6662.0	455.1
lower limit,	.05,	6	1061.0	7957.5	6662.6	454.9
maximum likelihood			904	7824	6920	488
upper limit,	.05,	7	766.8	7649.5	7196.3	523.4
upper limit,	.01,	7	740.5	7612.4	7246.2	536.9

Drosophila melanogaster X chromosome data of BRIDGES in MORGAN, BRIDGES and SCHULTZ (1935).

TABLE 2

A comparison of upper fiducial limits of strand and tetrad rank-frequencies calculated in two ways: A, highest tetrad rank same as highest strand; B, highest tetrad rank one higher than highest strand

Rank	0	1	2	3	4
Strand frequency					
P = .05 A	6456.1	7619.2	1995.4	65.4	0
B	6477.5	7607.8	1981.6	65.1	3.91
P = .01 A	6425.4	7630.7	2012.9	67.1	0
B	6446.1	7619.8	1999.2	66.2	4.61
Tetrad frequency					
P = .05 A	766.8	7649.5	7196.3	523.4	0
B	790.1	7648.6	7239.0	395.7	62.6
P = .01 A	740.5	7612.4	7246.2	536.9	0
B	763.8	7603.5	7312.4	382.6	73.7

Data of BRIDGES, in MORGAN, BRIDGES and SCHULTZ (1935).

the χ^2 for $(k + 1)$ degrees of freedom, rather than k; ii. to alter the right-hand side of (4), (6) and (7), respectively, to $(1 - \sigma_{k+1})$, $(G/N + 1)$ $(1 - \sigma_{k+1})$ and $[(G/N + 1)$ $(1 - \sigma_{k+1}) - 1]/V$. Once these changes have been made the procedure in evaluating $\sigma_0 \cdots \sigma_k$ is exactly as before. An example of the results, with the corresponding values for $\sigma_{k+1} = 0$, is shown in table 2.

SUMMARY

1. Although WEINSTEIN's equations (1936) give maximum likelihood estimates of the crossover-rank frequencies in the universe of tetrads from which a given sample of strands was derived, just as with every estimate of universe parameters from sample statistics other estimates within a certain range cannot be rejected.

2. Where four or more strand ranks are observed, the limits of the range form a closed shell of three or more dimensions containing, of course, infinite possible sets of tetrad-rank frequencies each of which could be regarded as a fiducial limit of the maximum likelihood estimate.

3. Among these infinite points three seem to be of principal interest:

(i.) that set of tetrad-rank frequencies for which (a) the χ^2 of the observed strand-rank frequencies would be some specified value, generally corresponding to P between .01 and .05, and (b) the average number of crossovers per tetrad is least;

(ii.) that set of tetrad-rank frequencies having (a) the same value of P and (b) the highest average number of crossovers per tetrad;

(iii.) like (ii) but including some tetrads one rank higher than the highest observed strand rank, in a proportion such that absence of strands of the same rank from the given sample would have some specified probability.

4. The lower and two upper fiducial limits, so defined, can be calculated by methods derived and illustrated here.

LITERATURE CITED

MORGAN, T. H., C. B. BRIDGES, and J. SCHULTZ, 1935   Report of investigations on the constitution of the germinal material in relation to heredity. Carnegie Inst. Wash. Ybk. **34:** 284–291.

WEINSTEIN, A., 1936   The theory of multiple strand crossing over. Genetics **21:** 155–199.

WHITTAKER, E. T., and G. ROBINSON, 1944   *The Calculus of Observations*. XVI + 397 pp. Blachie and Son, Ltd. London and Glasgow.