The American Journal of Human Genetics

Vol. 11 No. 4 *December 1959*

Some Comments on the Statistical Methods used in Linkage Investigations

CEDRIC A. B. SMITH

Galton Laboratory, University College London

NEWTON MORTON has shown in a series of papers (Morton 1955, 1957, Steinberg and Morton 1956) that sequential methods for the detection of linkage have many advantages. They are appreciably simpler and more efficient than most of the methods previously used, and are convenient in practice. Nevertheless their use raises important questions of principle. It seems to me that the use of a Wald sequential stop-rule, as advocated by Morton, is not appropriate in linkage work and confers no advantage: on the other hand Bayes' Theorem can be quite easily applied, using Morton's own results, and gives a more satisfactory answer from both the theoretical and practical points of view. (It may not in fact give a very different answer to that given by the sequential test: all the same it seems worth discussing the issues of principle involved.)

There are several lines of approach to the problems of statistical inference, namely (1) Bayes' Theorem, or inverse probability, (2) significance tests, (3) decision functions (4) sequential tests, (5) point estimation, including maximum likelihood, (6) interval estimation, including confidence and fiducial intervals. As these will be rather better known to the statistician than to the geneticist, it seems worth discussing the first four in some detail (for the others, see Kendall 1946).

(1) BAYES' THEOREM

I begin with a specific example. Consider a random mating population in which there are a pair of alleles G, g, each of frequency $\frac{1}{2}$, and with g recessive to G. By the Hardy-Weinberg rule the genotypes will occur in the ratio $\frac{1}{4}$ GG:$\frac{1}{2}$ Gg:$\frac{1}{4}$ gg. The intercross mating $Gg \times Gg$ will accordingly occur with frequency $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, i.e. one mating in four will be of this type. The backcross mating $Gg \times gg$ can be either Gg ♀ $\times$ gg ♂ or reciprocally Gg ♂ $\times$ gg ♀, each with frequency $\frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$; to simplify the argument we will not distinguish between reciprocal matings, and so the backcrosses will form a proportion $2 \times \frac{1}{8} = \frac{1}{4}$ of all matings.

Suppose now that we have a sibship of tested children, say, for definiteness, one dominant (G) younger sib and one recessive (g = gg) older sib, but with untested parents. We wish to say as much as possible about the mating type of the parents. We see at once that it must be either backcross or intercross, since other matings produce all dominant or all recessive children. The question which remains is accordingly to find the relative chances of it being of one type or the other.

It is clear that we need only consider families of two children, like the one observed: other sizes of family will be irrelevant. Among all families of two, $\frac{1}{4}$ will be

Received November 29, 1958.

intercrosses and ¼ backcrosses. Since intercrosses produce (in the long run) ¾ dominant (G) children, and ¼ recessive (g), the proportion of intercrosses which produce one younger G and one older g type sib will be ¾ × ¼: thus, out of all families of two, a fraction ¼ × ¾ × ¼ = 3/64 will be intercrosses producing a younger type G sib and an older g: call these families of type T_1. A similar argument shows that a fraction

$$\frac{1}{4} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16} = \frac{4}{64}$$

will be backcrosses producing a younger G and an older g, or, say, families of type T_2. Now apart from the phenotypes of the sibs we know nothing more about the observed family: we can accordingly imagine it as effectively chosen at random from among the whole group of families of types T_1 and T_2. Since the frequencies of these families are in the ratio $\frac{3}{64}T_1:\frac{4}{64}T_2$, i.e., three T_1 to every four T_2, it is evident that the observed family has chances in the ratio $\frac{3}{64}:\frac{4}{64} = 3:4$ of being type T_1 or T_2 respectively. The absolute probabilities of being T_1 and T_2 must accordingly be 3/7 and 4/7 respectively, since these are in the ratio 3:4 and sum to 1. Note that these can be calculated as

$$\frac{\frac{3}{64}}{\frac{3}{64} + \frac{4}{64}} = \frac{3}{7} \quad \text{and} \quad \frac{\frac{4}{64}}{\frac{3}{64} + \frac{4}{64}} = \frac{4}{7}$$

These probabilities have a quite definite meaning. If we select from the general population at random (or in an unbiased manner) all families with one younger dominant sib, and one older recessive, then a proportion 3/7 of them will be intercrosses, and a proportion 4/7 backcrosses. This is often called "inverse" probability, as it argues from an observed event to a previous one which has not been directly observed, or, as may be said, from "effect" to "cause." The general principle is as follows (Bayes 1763). Suppose that an event E can happen in k distinct ways W_1, W_2, ... W_k) (say, k possible "causes" of E). Suppose further that the way or cause W_r has a probability w_r of occurring in the general population, and that when W_r is known to have occurred, the event E has a probability e_r of following it. Then if we have observed that E has happened, but have no further knowledge about which way it happened, the probability that it did in fact happen through way W_r is

$$\text{Prob}\,(W_r|E) = w_r e_r/(w_1 e_1 + w_2 e_2 + \ldots + w_k e_k). \qquad (1)$$

The argument is essentially the same as in the genetical example, where the two "ways" W_1 and W_2 were the intercross and backcross mating respectively, with respective "initial" or "prior" probabilities $w_1 = ¼$, $w_2 = ¼$. The subsequent probabilities of getting the event E, i.e., the observed sibship, were $e_1 = 3/16$ from an intercross and $e_2 = ¼$ from a backcross. A substitution in (1) gives Prob $(W_1|E) =$ 3/7 as before.

To use formula (1) it is essential to know the initial probabilities w_r. This is made clear from a comparison of the following two situations:

(a) Matings of type G × g fall into two classes: either they are $GG \times gg$, and

produce only G children, or they are $Gg \times gg$, and produce equal numbers of G and g children. Suppose that a G × g mating taken at random gives rise to 10 dominant offspring: what is the chance that it is $GG \times gg$? If, as before, the gene frequencies of G and g are each $\frac{1}{2}$, the heterozygotes Gg are twice as frequent as the homozygotes GG, so the initial probabilities of the mating being $GG \times gg$ and $Gg \times gg$ are respectively $w_1 = \frac{1}{3}$ and $w_2 = \frac{2}{3}$ respectively. The respective probabilities of getting 10 dominants, given the mating, are $e_1 = 1$ and $e_2 = 1/2^{10} = 1/1024$. A substitution in (1) shows that the chance that the mating was in fact $GG \times gg$ is $w_1e_1/(w_1e_1 + w_2e_2) = 512/513$, i.e., near certainty, which is what common sense would suggest.

(b) A coin can (we assume) be either double-headed, in which case it will always come down heads, or normal, coming equally often heads and tails. We take a coin at random, toss it 10 times, and find it comes heads every time. What is the chance that it is double-headed? This is similar to the previous problem in every respect except the initial probabilities. We might perhaps know that one coin in a million is double-headed, so that now $w_1 = .000001$, $w_2 = .999999$: a calculation then would show that the chance that the coin is double-headed is only about .001. Even though the observed data are very unlikely to occur if the coin is normal, the alternative supposition that it is double-headed is still more unlikely, and so we rightly believe in the normality of the coin. In spite of the similarity of the two problems, the conclusions go in opposite directions.

(2) SIGNIFICANCE TESTS

A serious obstacle to the application of Bayes' Theorem in practice is that we often have only very vague ideas of what the initial probabilities are. Suppose, for example, we wish to know whether a given type of mating does or does not give a 1:1 segregation on the basis of the observed numbers from such matings. To decide this we would require to know the initial chance that a mating "chosen at random" gives a 1:1 segregation. Not only is this difficult to estimate, but it is also open to serious question whether the mating under study can be legitimately said to be "chosen at random." In consequence, Bayes' Theorem has fallen into disfavour and most statisticians now use "significance tests" or other alternative methods of inference.

The principle of a significance test is, in its essentials, that if we can show that the observed data are very unlikely to occur if a hypothesis H_0 is true, we can regard them as "significantly" opposed to the hypothesis, and in the absence of good reason to the contrary we will tend to disbelieve the hypothesis. Thus if the hypothesis is that a given mating should give a 1:1 segregation, the chance that it gives a family of 10 children all alike (i.e., all dominant or all recessive) is $2/2^{10} = 1/512$. If then such a family occurs, the hypothesis of a 1:1 segregation is "rejected at a significance level 1/512." This appears to be the point of view of Sir Ronald Fisher, who is responsible for many of the most important tests of significance: in his book (Fisher 1956, p. 39) he says:

"The force with which such a conclusion is supported is logically that of the simple disjunction: *Either* an exceptionally rare chance has occurred, *or* the theory ... is not true."

Significance tests have the convenient property that they summarize the data in

a unique numerical manner: two persons given the same data will find the same level of significance (or perhaps nearly the same, if slightly different methods of calculation are used). But on the other hand, the interpretation of the test will depend on the background, experience, and prejudices of the investigator: if he (rightly or wrongly) believes very strongly in the hypothesis tested (H_0), he will still continue to believe in it, even if it is contradicted by the data at a high level of significance. He will say that this apparent contradiction is due either to an exceptional chance fluctuation or to bad experimentation.

When Bayes' Theorem can be applied, it is more informative than a significance test, for it gives to each hypothesis an exact probability of being true. A significance test, on the other hand, may "reject a hypothesis at significance level P", but P here is *not* the probability that the hypothesis is true, and indeed the rejected hypothesis may still be probably true if the odds are sufficiently in its favour at the start. For example, in human genetics there are odds of the order of 22:1 in favour of two genes chosen at random being on different chromosomes; so even if a test indicates departure from independent segregation at the 5 per cent level of significance, this is not very strong evidence in favour of linkage. (Morton allows for this in his sequential tests.)

(3) DECISION PROCEDURES

An alternative theory of significance tests was given by Neyman & Pearson (1933) and this had developed into Wald's (1947, 1950) theory of decisions. A simple example is that of a manufacturer of articles who wishes to test a batch of them before sending them on to a purchaser. A certain small fraction may inevitably be expected to be defective: the manufacturer wishes to keep this fraction reasonably low without going to the expense of testing every article individually. He therefore takes a sample from the batch and tests it. If the sample is satisfactory he "accepts" the batch and sends it on, if not, he "rejects" the batch, retaining it for further tests or for destruction. Naturally, this procedure cannot be absolutely infallible; unless he takes the whole batch and tests each article individually, he can not know for certain exactly how good it is. (In some cases testing involves destroying the article and so such a complete test of a batch is in any case useless.) Among "good" batches, defined as those with only a certain small proportion of defectives (e.g. 1 per cent) there will be a few which will fail the test purely by chance, be rejected, and so involve the manufacturer in further expense, such as in testing the whole batch. Such a misleading rejection is called an "error of the first kind" and the proportion of good articles which are rejected in this way is usually denoted by α. Similarly among "bad" batches, (e.g. those with 5 per cent or more of defectives) a small proportion will pass the test by chance, and will cause trouble to the consumer. These are "errors of the second kind," and the proportion of bad batches which are misleadingly accepted is usually denoted by β. Both kinds of error are undesirable, so the general aim of the manufacturer will be to keep α and β as low as is possible without unreasonably increasing the cost of testing.

A test of this kind, which leads to a choice between two or more possible courses of action, is generally known as a "decision procedure".

(4) SEQUENTIAL TESTS

The simplest kind of decision procedure is the one considered above, in which batches have to be sorted into two categories, "good" and "bad." Wald (1947) showed that the most efficient way to proceed in such cases was not to take the whole sample at once and test it, but rather to draw items one by one until a sufficient number of good or bad articles have been drawn to make it clear which way the decision should go. Roughly speaking, this may be compared to the scoring system in tennis as contrasted with that in football. Tennis proceeds "sequentially," in that play goes on until one player has managed to get a definite lead of a certain number of points over the other (although the rules defining this lead are somewhat complicated in detail). The time needed to complete any match therefore depends on the relative strengths of the players; if they are nearly equal it may be some time before a final decision is reached, but such a decision indicates the better player with reasonable reliability. A football match is normally a "fixed sample size procedure," in that it goes on for a definite time and then ends. If the teams are very unequal, it may be clear which is the better one early in the game, whereas if they are nearly equally matched the final decision may be very much a matter of chance.

From the point of view of the manufacturer sequential tests have very much the same advantage of reaching a reliable conclusion as quickly as possible. The argument considered above suggests that this will be so if the "bad" batches are heterogeneous, for then by a sequential procedure the really bad batches will be detected very quickly, economizing in time and expense of testing, while the less bad ones will take longer, but no longer than is really necessary. A fact which is not quite so obvious, but which was demonstrated by Wald (1947), is that sequential tests are best even in the cases in which all "bad" batches are equally defective. Suppose that each good batch contains a known proportion p_0 of good articles, and $q_0 = 1 - p_0$ of defectives, whereas each bad batch contains a known proportion p_1 of good articles, q_1 of defectives. The method adopted is then as follows: for each article X we calculate a "score" $z(X) = \log (P_B/P_G)$, where P_G means the probability of getting an article like X when the batch is good, and P_B the same when the batch is bad, thus when X is good $P_G = p_0$ and $P_B = p_1$ and so

$$\left.\begin{array}{l}\text{if } X \text{ is good, } z(X) = \log (p_1/p_0) < 0;\\ \text{if } X \text{ is bad, } \quad z(X) = \log (q_1/q_0) > 0.\end{array}\right\} \tag{2}$$

As the articles are drawn from the batch, their sources are added consecutively to make a running total score Z. We also choose two "goals," a positive number a and a negative number b. If at any time the total score Z reaches a, we decide to reject the batch, whereas if it falls to b, the batch is accepted. If we define as before α to be the proportion rejected out of all good batches, and β the proportion accepted of all bad batches then nearly enough $a = -\log \alpha$ and $b = \log \beta$ when α and β are sufficiently small. Thus one can design a test with predetermined values of α and β. [More accurately $a = \log (1 - \beta) - \log \alpha$ and $b = \log \beta - \log (1 - \alpha)$].

In practice one rarely meets with quite so simple a situation in which all "bad" batches are equally bad. The usual custom is accordingly to proceed roughly as

follows. Let q_0 be the proportion of defectives produced when the process is working well, i.e. in good batches. Let q_1 be the maximum proportion of defectives that the manufacturer is prepared to tolerate, i.e. any batch worse than this ought to be rejected. Write $p_0 = 1 - q_0$, $p_1 = 1 - q_1$, as before. Then we proceed with a sequential test for comparison between proportions q_0 and q_1 of defectives: the scores are calculated from (2) and the addition of scores proceeds as before until one of the goals a or b is reached. This test accordingly has the following properties:

(i) If the proportion of defectives is q_0, the batch is good: from the preceding theory we then know that the batch has a chance α of being mistakenly rejected, and $(1 - \alpha)$ of being accepted. Now in a significance test the "significance level" is defined as the probability of rejecting the hypothesis H_0 tested, when H_0 is in fact true; so by analogy it is natural to refer to α here as the "significance level" of the sequential test, i.e. the probability of mistakenly rejecting a batch when it is in fact good.

(ii) If the proportion of defectives is in fact q_1, the chance of accepting the batch is β and of rejecting it is $(1 - \beta)$. Since β is ordinarily chosen to be a small number, the majority of such batches will be rejected.

(iii) If the proportion of defectives is greater than q_1, the batch is definitely bad. Since each defective scores positively the score will normally tend to the positive goal a even more rapidly than if the proportion was equal to q_1; so on the average the batch is rejected even more quickly, and the chance of it being mistakenly accepted is even smaller. Thus β is an upper bound to the proportion among all bad batches of those which are mistakenly accepted ("errors of the second kind").

(iv) If the proportion of defectives is between q_0 and q_1 the batch is tolerable, even if not completely good. In such a case the manufacturer will not be over much concerned whether the batch is accepted or rejected.

This procedure seems therefore to give very reasonable results. The test used will be the one best adapted to distinguish between the two proportions q_0 and q_1 of defectives: it will therefore not be absolutely the best test for the real situation in which we get varying proportions of defectives in the bad batches. However, the computations in this simplified test are so easy that it is usual to ignore this slight loss of efficiency in the interests of ease of calculation. One could devise a sequential test which would take into account the varability of the proportion of defectives, but it would be rather more cumbersome, and we will not discuss it here. The same applies to a "closed sequential test" in which the sampling proceeds sequentially up to a certain fixed size, and then stops.

SEQUENTIAL TESTS APPLIED TO LINKAGE PROBLEMS

Morton (1955) has proposed that sequential tests should be applied to the detection of linkage. This can very readily be done by a verbal adaptation of the preceding argument. Instead of drawing a sample of individuals from a batch of goods, we collect families from a population, and study a particular pair of (autosomal) loci. Instead of asking whether the batch is "good" or "bad" we ask whether the loci are respectively unlinked or linked. Instead of saying that some batches are so nearly good as to be tolerable, Morton argues that some pairs of loci are so loosely linked that it would not be practicable to try to detect the linkage in any reasonably

attainable body of data. We therefore choose a value θ_1 of the recombination fraction which represents the loosest linkage we can reasonably expect to detect, and compare this with no linkage, i.e. recombination fraction $\frac{1}{2}$. The score for any family F is defined to be

$$Z(F) = \log \text{Prob}\ (F|\theta_1) - \log \text{Prob}\ (F|\tfrac{1}{2}) \tag{3}$$

where Prob $(F|\theta_1)$ is the probability of occurrence of the family F when the recombination fraction is θ_1, and Prob $(F|\frac{1}{2})$ is the similar probability in the absence of linkage. These scores for separate families are added successively until the total score reaches either the positive goal a, indicating the presence of linkage, or the negative goal b, indicating its absence. For details of the method of calculation of scores, and tabulations of their values in various commonly occurring types of family see Morton (1955, 1957).

THE PRINCIPLE OF SELF-SUFFICIENCY

Suppose we have taken at random a sample S of individuals from a population, and wish to consider what information this gives about the population. To take a simple example, suppose we are interested in the proportion q of defectives, and that the sample contains n individuals, of which x were found to be normal and y defective. The information given by the sample may be considered to have three aspects.

(i) The numbers x, y of normals and defectives in the sample;

(ii) The order in which the normals and defectives occur (e.g.) NDDND or DNNDD etc.);

(iii) The fact that the sampling stopped at that point, and did not continue further.

Now once we are given the numbers x, y of normals and defectives respectively, the order in which they occur is purely a matter of chance. Some arrangements may be ruled out by the stopping rule, in that if such an order of normals and defectives has occurred, the sampling would have been brought to an end at some previous stage. But all arrangements which are not ruled out in this way are equally likely to occur, and so the particular order observed gives no information about the proportion q of defectives (though it may give information about the stopping rule). As for the fact that sampling has been stopped at this particular point, it seems clear that this can contribute no additional information about q, since it is essentially a decision entirely at the discretion of the investigator. If he bases it on some stop rule which depends on the number or order of occurrence of normals and defectives, this merely uses the information (i) and (ii) over again; if he bases it on any other considerations it will be irrelevant to the issue. (Any extra-sensory perception of the correct answer would have to be considered as further sampling, whereas we assume that sampling has stopped!) This intuitive kind of argument can be extended to more complicated cases, and shows that so long as sampling is at random, the information which any sample provides about the population from which it is drawn is contained entirely in the numbers of the different types of objects occurring in the sample. The order in which they occur, and the stop rule which ends the sampling, are both completely irrelevant. We might call this the "Principles of Self-sufficiency of the Sample."

It seems a reasonable deduction from this that the statistical analysis of the results should depend only on the numbers observed, and not on the order of occurrence or on the stopping rule. It does not matter whether the experimenter has finished because of a sequential, closed sequential or fixed size stop rule, or because he has exhausted all available data, or because he is just bored, the conclusions to be drawn from his data should be the same, otherwise the statistician is being inconsistent, in drawing sometimes one conclusion and sometimes another from the same information.

These conclusions are in harmony with various investigations: for example, Anscombe (1953, 1954) in papers on sequential estimation, points out that estimators used in ordinary statistical theory for samples of fixed size will apply equally well to large samples obtained sequentially: e.g. the estimate of the proportion q of defectives will still be y/n, with standard error $\sqrt{(pq/n)}$ (where $p = 1 - q$) as usual. Barnard (1953) points out that the maximum likelihood estimates do not depend on the stop rule. Wald (1950) shows that the "best" decision procedures must be of the so-called "Bayes" type, which means that they are calculated from the likelihood function for the sample, which does not depend on the stop rule. (In technical terminology, the Bayes decisions form a complete class of admissible decision functions.)

Unfortunately, if this argument is accepted it leads to some rather startling conclusions. Very few statistical procedures in general use at present are valid independently of the stop rule. Significance tests are a particularly awkward exception; if the principle of self-sufficiency of the sample was to apply to them, it would mean that we should be able to apply in all cases the ordinary significance tests designed for a sample of a fixed size. However, it is very well known that one can often cheat (intentionally or otherwise) by continuing to gather data until a particularly favourable chance fluctuation occurs, and then stopping and applying the usual significance test to show that this fluctuation is "significant". Anscombe (1954) shows how to "prove" or "disprove" any given statistical hypothesis in this way. Of course, this does not mean that ordinary statistical procedures do not work reasonably well when used properly; but it does suggest rather strongly that they are not entirely satisfactory, and it is conceivable that in due course some new methods may be put forward which do not have these disadvantages, or lead to these paradoxical results in certain situations, and which will gain general acceptance. I shall show later, however, that very fortunately in the particular case of linkage detection the difficulties can be overcome without introducing any new or unconventional methods.

What precisely is the advantage of a sequential procedure? It cannot extract any more information from a given sample than can fixed-size sampling, for as we have seen, the information depends only on the sample and not on the stop rule. It seems therefore that the advantage is simply that of finishing sampling when we have collected sufficient data for the purpose in hand. If the object is to accept or reject a batch of goods, as in the industrial case already considered, Wald's stop rule may be the appropriate one. But it is not difficult to think of cases where other stop rules would be more appropriate. For example, the proportion q of defectives in the sample will be estimated as y/n, and the standard error of this estimate will be

nearly enough $\sqrt{(pq/n)} \simeq \sqrt{(xy/n^3)}$. If therefore we wish to estimate the proportion q with some definite accuracy, i.e. with a standard error not exceeding some given value S_0, we must continue collecting data until $\sqrt{(xy/n^3)} < S_0$, but will then presumably stop as the required accuracy has been attained.

COMMENTS ON MORTON'S SEQUENTIAL LINKAGE TEST

In the case of an industrial experiment, the object is to make a choice between two definite courses of action, namely the acceptance or rejection of a batch of goods. In translating this into a sequential linkage test, the corresponding object becomes verbally to "accept" or "reject" the hypothesis of linkage. These phrases "accepting" or "rejecting a hypothesis" are much used in the literature of significance test, but the question arises as to what they mean, if anything. If we "accept" or "reject" a hypothesis, we do not in general perform any actions as a direct consequence of the decision. (Such more remote problems as giving eugenic advice will be considered later.) Nor does "acceptance" mean unconditional and complete belief in the existence of linkage, since any sample can only make linkage seem more or less probable, never completely certain. The situation is that any body of data will leave us more or less inclined to believe in linkage between the loci under study, and the reasonable procedure would be to construct some measure of how convincing the data are. But this is not a decision problem in the technical sense of the word; it does not call for a definite choice between two or more alternative courses of action, and accordingly there does not seem any particular point in using a decision procedure, or a Wald sequential test. It is not easy to think of any situations in the real world which would correspond to "acceptance" or "rejection," in the sense that these are precisely defined decisions, and that the object of the investigation has been fulfilled when one or the other has happened. We could perhaps think of "acceptance" as being sufficiently strongly convinced of the existence of linkage, and "rejection" as sufficiently strong conviction of its absence, and so reduce the question formally to a decision problem. However, the operation of a sequential stop rule would then mean that one was only interested in reaching that degree of conviction; when it has been reached, one would stop sampling, and refuse to consider further evidence, turning instead to some other problem. This would be very odd behaviour if the data favored linkage; one would rather expect the investigator to gather fresh data with increased eagerness. But the Wald sequential procedure does not make any provision for continuing after the "stop." This would be rather like getting two tennis players to play a match, decide who had won, and then continue playing for some extra time to confirm the decision. That is not efficient, for if they have to play this extra time, it would be more reliable to decide who had won on the basis of the final results than on the less complete results at the intermediate point. (We assume that there is no variation in the strength of the players as a result of fatigue or practice effects, since in the analogous statistical tests the probabilities of "defective" or "linkage" are supposed to remain constant.)

In view of the above discussion it seems best to divide the questions to be answered in two parts.

(i) What information does the sample give us about linkage?

(ii) Is it worth while continuing to gather further data?

In a Wald sequential test these questions are bound together; at the same time as a decision is made in favor of or against linkage, it is also decided to stop sampling. Nevertheless these two questions are somewhat different in nature: the second one is a decision problem, calling for a choice between the possibilities of continuing the investigation or abandoning it. The first is not a decision problem, and is not really appropriately dealt with by a sequential test.

BAYES' THEOREM APPLIED TO LINKAGE ESTIMATION

Since our argument suggests that neither significance tests nor sequential tests are entirely appropriate for the analysis of linkage data, the question arises urgently whether there is any accepted statistical method which will do better. Fortunately it is possible to apply Bayes' Theorem. The basis of this is Morton's (1955) demonstration, on theoretical and empirical grounds, that we can give a reasonaby good approximation to the initial distribution of the recombination fraction θ. Since there are 22 pairs of autosomes in man, there is a probability of about $^{21}/_{22}$ that two autosomal genes chosen at random will be on different chromosomes; in the remaining case of linkage, the recombination fraction will have an approximately uniform probability of taking any value between 0 and $\frac{1}{2}$. The situation is directly comparable with that considered in equation (1), except that instead of finding the probability that a given event E happened through way W_r, we have to find the probability that a given sample S arose in a population with recombination fraction θ. There is an additional complication in that instead of a finite number of ways W_r we have a continuous range of values of θ. This last difficulty can be overcome with negligible loss of accuracy by taking only the first two decimal places in θ i.e. by proceeding as if θ could take only the exact values .00, .01, .02, ... .50. The initial probability $^{1}/_{22}$ of their being linkage must be divided equally between the 50 values from $\theta = .00$ to $\theta = .49$, giving each one an initial probability $^{1}/_{1100}$; the remaining value $\theta = .50$, corresponding to no linkage, has initial probability $^{21}/_{22}$. Let $e(\theta) =$ Prob $(S|\theta)$ denote the chance of the observed sample S arising when the recombination fraction is known to be θ; by (1) the choice that the recombination fraction is θ after the observation is S is, when $\theta < .50$

$$\text{Prob } (\theta \mid S) \frac{\frac{1}{1100} e(\theta)}{\frac{1}{1100}[e(.00) + e(.01) + \cdots + e(.49)] + \frac{3}{2}\frac{1}{2} e(.50)} \tag{4}$$

The value of Prob $(\theta = .5|S)$ has the same denominator, but the numerator is now $\frac{3}{2}\frac{1}{2}e(.50)$. Now we can simplify (4) in the following way. Write

$$\lambda(\theta) = e(\theta)/e(.50) = \text{Prob } (S|\theta)/\text{Prob } (S|.50) \tag{5}$$

Furthermore, let Λ denote the average value of $\lambda(\theta)$, values of θ lying between 0 and .49, i.e.

$$\Lambda = [\lambda(.00) + \lambda(.01) + \ldots. + \lambda(.49)]/50$$

Then if we multiply numerator and denominator on the right hand side of (4) by $50/e(.50)$ it becomes

$$\text{Prob } (\theta|S) = \lambda(\theta)/50\ [\Lambda + 21] \tag{6}$$

when $\theta < .50$. This probability means, of course, the chance that the recombination fraction takes the value θ *correct to two decimal places*. In the same way, we find that the probability that the loci are unlinked is

$$\text{Prob } (\theta = .50|S) = 21/[\Lambda + 21] \tag{7}$$

In order to use these formulas it only remains to find the values of $\lambda(\theta)$ and Λ. Now we can show that

$$\lambda(\theta) = \text{antilog } Z(\theta) \tag{8}$$

where $Z(\theta)$ is the total lod score. Suppose that the sample consists of families F_1, $F_2 \ldots . F_h$. By the multiplication law of probabilities, the probability of occurrence of the whole sample, given θ is

$$\text{Prob } (S|\theta) = \text{Prob } (F_1|\theta)\ \text{Prob } (F_2|\theta) \ldots \text{Prob } (F_h|\theta)$$

Therefore

$$\lambda\ (\theta) = \frac{\text{Prob } (S \mid \theta)}{\text{Prob } (S \mid \tfrac{1}{2})} = \left(\frac{\text{Prob } F_1 \mid \theta}{\text{Prob } F_1 \mid \tfrac{1}{2}}\right) \cdots \left(\frac{\text{Prob } (F_h \mid \theta}{\text{Prob } (F_h \mid \tfrac{1}{2}}\right) \tag{9}$$

But by definition, the score $z_r(\theta)$ for family F_r is

$$z_r\ (\theta) = \log \text{Prob } (F_r \mid \theta) - \log \text{Prob } (F_r \mid \tfrac{1}{2}) = \log \frac{\text{Prob } (F_r \mid \theta)}{\text{Prob } (F_r \mid \tfrac{1}{2})}$$

Hence by taking logarithms on both sides of (9) we get

$$\log \lambda(\theta) = z_1(\theta) + z_2(\theta) + \ldots + z_h(\theta) = Z(\theta)$$

where $Z(\theta)$ means the total score, being the sum of the scores for each family. This is equivalent to (8). Nor ordinarily it is tiresome to calculate the exact score $z_r(\theta)$ for every family, taking into account the whole of the available information. But as Morton points out, the calculations can be very greatly simplified with only a small loss of information by including those families in which the genotypes of the parents are completely known (except for phase), and applying a correction to their scores to allow for the omission of the other families. In this case we can find the values of $z(\theta)$, for $\theta = .05, .1, .2, .3$ and $.4$ from Morton's (1955, 1957) tables. For $\theta = .5$, $Z(\theta)$

TABLE 1. STEINBERG AND MORTON'S SCORES

Recombination fraction θ	0	.05	.1	.2	.3	.4	5
Total score $Z(\theta)$	$-\infty$	−2.631	−.274	.716	.528	.175	0
$\lambda(\theta)$ = antilog $Z(\theta)$	0	.002	.532	5.200	3.373	1.496	1

is always zero by definition. The value of $Z(0)$ is not given, but in any case in which there is at least one certain crossover (as will happen in the great majority of linkage investigations) the total score Z(0) is $-\infty$, and $\lambda(0) = 0$. For example, the total scores found by Steinberg and Morton (1956) for linkage between cystic fibrosis and the MNS blood groups are shown in Table 1. This only gives a few values of $\lambda(\theta)$. Although a completely accurate determination of $\lambda(\theta)$ for other values of θ would require an extension of Morton's tables, a great deal can be done without this. Λ is by definition the average value of $\lambda(\theta)$ for values of θ between 0 and $\frac{1}{2}$; this can be calculated approximately by a formula for numerical integration, such as Simpson's rule.

$$\Lambda \simeq \tfrac{1}{30}\,[\lambda(0) + 4\lambda(.05) + 3\lambda(.1) + 8\lambda(.2) + 4\lambda(.3) + 8\lambda(.4) + 2] \qquad (10)$$

This gives approximately $\Lambda = 2.355$; whence from (7) the probability that the genes are not linked is Prob $(\theta = \frac{1}{2}|S) = 21/(\Lambda + 21) = .899$. However, it is rather more accurate to use interpolation formulas to find intermediate values of $\lambda(\theta)$. The following formulas, obtained by fitting a polynomial to $Z(\theta)$, are convenient and of high accuracy.

$$\begin{aligned}
Z(.15) &= -.046Z(.05) + .345Z(.1) + 1.203Z(.2) - 1.148Z(.3) + 1.422Z(.4)\\
Z(.25) &= .007Z(.05) - .033Z(.1) + .355Z(.2) + 1.053Z(.3) - .762Z(.4)\\
Z(.35) &= -.002Z(.05) + .007Z(.1) - .\,043Z(.2) + .402Z(.3) + .949Z(.4)\\
Z(.45) &= - .001Z(.1) + .008Z(.2) - .\,044Z(.3) + .373Z(.4)
\end{aligned}$$

Using these we get from Table 1 the following interpolated values:

Recombination fraction θ	.15	.25	.35	.45
Interpolated score $Z(\theta)$	.531	.667	.351	.048
$\lambda(\theta)$	3.396	4.645	2.244	1.117

By using these interpolated values of $\lambda(\theta)$ we can calculate Λ from Simpson's rule.

$$\begin{aligned}
\Lambda \simeq \tfrac{1}{30}\,[&\lambda(0) + 4\lambda(.05) + 2\lambda(.1) + 4\lambda(.15) + 2\lambda(.2) + 4\lambda(.25)\\
&+ 2\lambda(.3) + 4\lambda(.35) + 2\lambda(.4) + 4\lambda(.45) + 1]
\end{aligned} \qquad (11)$$

Thus, on substituting the values calculated above we find $\Lambda = 2.261$ whence a more accurate estimate of the probability of the loci being unlinked is $21/\Lambda + 21) = .889$. For $\theta < .5$) we find from (6) the values of Prob $(\theta|S)$ in Table 2. These values are shown in Fig. 1. (It should be remembered that Prob $(\theta|S)$ means here the probability of the recombination fraction θ correct to two places, i.e. Prob $(\theta = .20|S)$ is the chance that θ lies between .20 and .21). Other values could be obtained if desired by interpolation, but these are sufficient to show that the distribution has a peak around $\theta = .2$.

TABLE 2. PROBABILITY OF VARIOUS VALUES OF THE RECOMBINATION FRACTION

θ	.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.49
Prob $(\theta\,\vert\,S)$	0	.00000	.00046	.00292	.00447	.00399	.00290	.00193	.00129	.00096	.00086

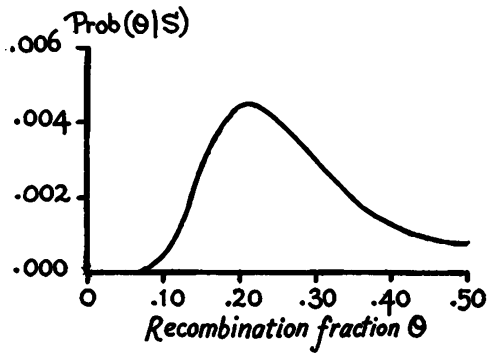

FIG. 1. Probabilities $P(\theta|S)$ of the occurrence of values of the recombination fraction θ less than .5. (There is a concentration of probability at $\theta = .5$, corresponding to the absence of linkage).

ADVANTAGES OF BAYES' THEOREM

Bayes' Theorem has the advantage that, unlike any other method of statistical inference, it gives the answer directly as a probability. Thus Steinberg and Morton's data gives a probability .9 of their being no linkage, and only .1 of linkage being present. This probability has a direct meaning, and does not need to be hedged about with qualifications, unlike a significance level. The investigator can stop sampling at any point he pleases, and for any motive, and provided he calculates the probabilities from all the data he has gathered up to that point, they will be valid. If we consider the linkage between one main character and a variety of test characters, such as blood groups, PTC taste testing, etc., it is well known that on the basis of testing by significance we can expect one test in 20 to give a significance at the 5 per cent level, or one in 100 at the 1 per cent level. The experimenter is accordingly warned not to take an occasional significance too seriously, and not to pick out the more hopeful cases and test them alone for significance. However, no such special precautions are needed in using final probabilities: each linkage test gives its own result, and can be considered entirely on its own without reference to the others. Bayes' Theorem also has the advantage of giving not only the chance of linkage being present but also the probabilities of different values of θ between 0 and .5, i.e. it combines the "detection" and "estimation" of linkage in a single calculation. With the help of Morton's (1955, 1957) tables, the amount of work required to do this is quite small—much smaller than the effort required to gather the data in the first place.

However, it must be remembered that all these useful properties depend on the availability of a valid initial distribution for θ. The one given by Morton, although very plausible, can only be approximate, and hence the probabilities calculated from it will be similarly approximate. Such a degree of vagueness seems inescapable at present; if the initial probabilities are doubtful to this extent, so also must be the final probabilities. Another source of difficulty, which will be resolved only by further

investigations, is that there may be a difference between the distributions of θ in the male and female. A certain caution is therefore necessary in interpreting the results.

Note. The interpolation formulas given above for calculating the values of $Z(.15)$, $Z(.25)$, $Z(.35)$, $Z(.45)$ were obtained on the assumption that all families were of two generations at most, with parental phases unknown. When that is not so, the formulas may be wrong and the following ones should be used instead.

$$
\begin{aligned}
Z(.15) &= -.111Z(.05) + .547Z(.1) + .729Z(.2) - .219Z(.3) + .062Z(.4) \\
Z(.25) &= .048Z(.05) - .156Z(.1) + .625Z(.2) + .562Z(.3) - .089Z(.4) \\
Z(.35) &= -.048(.05) + .141Z(.1) - .312Z(.2) + .834Z(.3) + .042Z(.4) \\
Z(.45) &= .111Z(.05) - .312Z(.1) + .583Z(.2) - .875Z(.3) + 1.250Z(.4)
\end{aligned}
$$

THE SAMPLING STOP RULE

As we have seen, when Bayes' Theorem is used, the investigator is not under any compulsion either to continue collecting data or to stop at any point merely in order to make the statistical analysis possible and valid. Nevertheless he will wish to know roughly how much data he must collect in order to have a reasonable chance of detecting linkage, and whether, after some data have been collected, it is worth while proceeding further.

Morton in his 1955 paper discussed how many families would be required to show the presence or absence of linkage according to his sequential tests. Although his methods differ in principle from ours, they will nevertheless indicate reasonably closely the order of magnitude of the number of observations required to raise the odds in favour of linkage to a reasonable value (such as 20:1); from his Fig. 4 it seems that we might reasonably need the equivalent of between 30 and 1000 double-backcross sib-pairs to do this, according to circumstances. With much less than this amount of data it scarcely seems worth beginning the calculations as the results will almost certainly be quite inconclusive.

Once the investigation has been begun, the question may arise of whether to continue it further or to abandon it. In principle this is a decision problem, and could be treated exactly by the methods of decision function theory However, there are so many complicating factors, such as the availability of data, the interest of the problem from aspects other than linkage, etc., that it seems simpler to treat the problem much more informally. The two main factors which guide the experimenter in making his decision whether to continue are (a) the chance of ultimate success (b) the probable amount of further data required to reach this success. However the chance (a) of ultimate success can be directly measured; it is the probability that the loci are linked; e.g. in Steinberg and Morton's data it is about .1. This in itself is better than the initial chance 1/22 before any data is gathered, but it is still scarcely encouraging. It is much harder to give any precise estimate of (b), because of sampling fluctuations. However, one can get a rough idea in the following way. Since the total score Z is the sum of the scores for the separate families, it will be proportional to the number of families, apart from sampling fluctuations. Thus if 3 times as much data was gathered on the linkage between cystic fibrosis and MNS, we might expect that

the Z values of Steinberg and Morton would be approximately multiplied by 3, i.e. they would become as follows:

$\theta =$	.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
$Z =$	-00	-9.89	$-.82$	1.59	2.15	2.00	1.58	1.05	.52	.14	0

On this basis the probability of the existence of linkage would be found as above to be $\Lambda/(\Lambda + 21) = .61$, i.e. if the trend continues, it will then be more likely than not that linkage is present. Similarly, with scores multiplied by 5 the probability of linkage would become .96, i.e. reasonably near certainty. The situation can therefore be summarized in the statements that there is only about 1 chance in 10 of linkage being found, and that if it is present it would probably require something like 5 times the present amount of data to demonstrate it with a reasonably high probability. The investigator will be able to decide on the basis of this information whether he feels it worth continuing.

EUGENIC ADVICE

Eugenic advice can in principle be of two kinds. It may be a simple statement of the probability of a certain defect developing, or the probability that if children were born from a given mating, they would have some defect. This is an application of the calculus of probabilities, taken in conjunction with Mendelian ratios, distributions of ages of onset, maternal age effects, etc., as far as these are relevant. Alternatively it could give a straightforward recommendation not to have children in certain cases; this is of the nature of a decision problem. However, it seems best in general to confine oneself to a calculation of the probabilities, and to allow the persons involved to make the decision for themselves, in the light of all relevant circumstances.

Sometimes such a calculation depends on the value of a recombination fraction θ, when the genes responsible for the defect are linked to some markers, such as blood groups. If the value of θ is accurately known, the calculation is usually straightforward: we find a probability Prob $(D|\theta) = d(\theta)$ say, of a certain "defect" D occurring. If however the value of θ is as yet uncertain, we calculate from all available data S a probability Prob $(\theta|S) = p(\theta)$ say, that the recombination fraction takes a given value θ. (We can again suppose for convenience that θ is given to two decimal places only.) The total probability of the defect occurring in the light of all available evidence, is then

$$\text{Prob } D = \Sigma p(\theta)\, d(\theta) \tag{12}$$

Where the summation is over all values of θ. However, it would be tedious to calculate p(θ) for every θ from .00 to .50; as a rule we will find it only for $\theta = .00, .05, .10 \ldots .50$ only, by intervals of .05. The value of Prob D is then given approximately by the following formula based on Simpson's rule:

$$\begin{aligned}\text{Prob } D = \tfrac{5}{3}\,[&p(.00)d(.00) + 4p(.05)d(.05) + 2(.10)d(.10) + 4p(.15)d(.15)\\ &+ 2p(.20)d(.20) + 4p(.25)d(.25) + 2p(.30)d(.30) + 4p(.35)d(.35)\\ &+ 2p(.40)d(.40) + 4p(.45)d(.45) + .6010p(.50)d(.50)]\end{aligned} \tag{13}$$

As a simple example, consider the position of a couple A ♂ × B ♀ who are contemplating having a child. A's brother E is married to B's sister F and they already have a child with cystic fibrosis, *cf cf*. If A, B, E and F are all of blood group M, while all their parents have group MN what is the probability that the child of A and B will also have cystic fibrosis?

Let θ denote the recombination fraction between MN and cystic fibrosis (with $\theta = \frac{1}{2}$ when they are not linked). E and F must both have genotypes MCf/Mcf. Since *cf* is a rare gene, A can be expected to be of genotype *Cf cf* (i.e. MCf/Mcf) either if the parents of A and E are $Mcf/NCf \times MCf/NCf$ and there are two non-crossovers or if the parents are $MCf/Ncf \times MCf/NCf$, and there are two crossovers. The chance that A is *Cfcf* is accordingly $(1 - \theta^2) + \theta^2$; the chance that B is *Cfcf* is similarly $(1 - \theta)^2 + \theta^2$. The child will have cystic fibrosis only if both parents are *Cfcf*, and then only with probability $\frac{1}{4}$; hence the chance of a defective child D is $d(\theta) = \frac{1}{4}[(1 - \theta)^2 + \theta^2]^2$. By working out the values of $d(\theta)$ for $\theta = .00, .05, \ldots .50$, and using the values of $p(\theta) = \text{Prob}\ (\theta|S)$ already obtained from Steinberg and Morton's data (Table 2, together with $p(.50) = .889$) we find from (13) that Prob $D = .0655$. This compares with the probability $1/16 = .0625$ which would be found without taking the possible linkage into consideration. It seems that in this case, where it is doubtful whether linkage exists, the probability is scarcely altered by taking it into account. If we knew that there was close linkage, Prob D might rise as high as .25.

SUMMARY

By using Bayes' Theorem together with the initial distribution for recombination fractions established by Morton it is possible to present the results of linkage tests in probability form. This combines detection and estimation of linkage, and can be used directly for further calculation. By the use of Morton's tables of lod scores the calculations can be done very quickly, and the process has a number of advantages over sequential tests.

REFERENCES

ANSCOMBE, F. J. 1953. Sequential estimation. *J. R. Statist. Soc., Lond.* B. 15: 1–21.

ANSCOMBE, F. J. 1954. Fixed sample size analysis of sequential observations. *Biometrics* 10: 89–100.

BARNARD, G. A. 1953. (Discussion on Anscombe's 1953 paper). *J. R. Statist. Soc., Lond.* B. 15: 22–23.

BAYES, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Tr. R. Soc., Lond.* 53: 370–418.

FISHER, R. A. 1956. *Statistical methods and scientific inference.* Edinburgh: Oliver and Boyd.

KENDALL, M. G. 1946. *The advanced theory of statistics. vol. 2.* London: Griffin.

MORTON, N. E. 1955. Sequential tests for the detection of linkage. *Am. J. Human. Genet.* 7: 277–318.

MORTON, N. E. 1957. Further scoring types in sequential linkage tests with a critical review of autosomal and partial sex-linkage in man. *Am. J. Human. Genet.* 9: 55–75.

NEYMAN, J., AND E. S. PEARSON. 1953. On the problem of the most efficient test of statistical hypotheses. *Philos. Tr. R. Soc. Lond.* 231: 289–337.

STEINBERG, A. G., AND N. E. MORTON. 1956. Sequential test for linkage between cystic fibrosis of the pancreas and the *MNS* locus. *Am. J. Human. Genet.* 8: 177–189.

WALD, A. 1947. *Sequential analysis.* New York: Wiley.

WALD, A. 1950. *Statistical decision functions.* New York: Wiley.