THE RANDOM PHENOTYPE CONCEPT, WITH APPLICATIONS¹

ALLAN BIRNBAUM

Courant Institute of Mathematical Sciences, New York University, New York, N.Y. 10012

Manuscript received January 12, 1972

Revised copy received August 21, 1972

ABSTRACT

A random phenotype is defined as a probability distribution over any given set of phenotypes. This includes as special cases the kinds of phenotypes usually considered (qualitative, quantitative, and threshold characters) and all others. Correspondingly general methods are indicated for analyzing data of all forms in terms of the classical Mendelian factor concept (as distinct from the biometrical methods usually applied to measurement and graded data, associated with the effective factor concept). These are applied in a new analysis of the data of E. L. GREEN (1951, 1954, 1962) on skeletal variation in the mouse. The adequacies of various classical one-factor and several-factor models are considered. Indications of an underlying scale are found from this new standpoint. The results are compared with those obtained by GREEN using the scaling approach. An illustrative application is also made to some of BRUELL's (1962) continuous behavioural data on mice. This work was substantially completed in 1959 but not previously prepared for publication. The same approach was originated and developed independently by R. L. COLLINS who has treated a wider range of theoretical problems (cf. COLLINS 1967, 1968a, 1969b, 1970c) and a wider range of applications (cf. COLLINS and FULLER 1968; COLLINS 1968b, 1969a, 1970a). A less general independent development is that of MODE and GASSER 1972.

IT is not generally appreciated that the classical genetic concepts introduced by MENDEL are directly and simply applicable to phenotypic data of all forms. This possibility has been ignored particularly when data represent each individual just by a numerical measurement. The terms "continuous variation" and "quantitative inheritance" are in standard use both to characterize all such genetic data, and to identify the only generally known methods of genetic interpretation of such data, the methods whose central concept is the *effective factor*, as distinct from the *classical factor* concept formulated by MENDEL (cf. e.g. MATHER and JINKS 1971). However with such data, and indeed with data of all possible forms (including for example a multivariate observation on each individual), once the question of an analysis in terms of classical factors is formulated clearly, it is not difficult to see how to implement such analysis by applying familiar Mendelian concepts. The general form of such analysis is given in Section 2; this is applied in Sections 3 and 4 to data published previously with analyses in terms of effective factors. The reader may turn to those sections next if he wishes to defer reading the following general background comments.

¹ Work supported in part by the U.S. Public Health Service (Grant No. 5 RO1 GM16202-03), Office of Naval Research, the National Science Foundation, and the Guggenheim Foundation.

Section 1. Concepts: The concept of a *random phenotype* is defined as follows:

(1) the phenotypes considered may be *any* chosen set of alternative partial descriptions of an individual's possible life histories (not necessarily either real-valued measurements or else discrete categories); and

(2) the observable expression of any given genotype (under given conditions of development and observation) may be any probability distribution over the indicated set of alternative descriptions.

Clearly neither *part* of this definition is new: Part (1) reflects the standard genetic concept of an individual's *norm of reaction*, which may be traced to the very general concept of the phenotype as introduced by JOHANNSEN (e.g. 1911, pp. 134–135). Part (2) reflects the generality which is standard in modern work in mathematical statistics and probability, allowing virtually any samples spaces, and any (properly defined) distributions. However the natural synthesis of these is a concept whose significance has not been generally appreciated.

With all possible forms of genetic data and distributions, the familiar laws of Mendelian genetics (including cases of linkage) find empirical expression as certain *linear restrictions* among the probability functions representing the random phenotypes of respective genotypes. The case of non-random (or certain) determination of a phenotype by a genotype is included as a special case of these formulations. This is quite compatible with special interpretations which may be given to certain non-random phenotypes on theoretical grounds (e.g. morphological or biochemical traits). The characterization of a non-random phenotype usually involves conventional as well as empirical elements, since non-random phenotypes will often be replaced by random ones when phenotypic descriptions are multiplied by introduction of finer distinctions, whether or not these may have genetic significance. From a standpoint of general theory it seems appropriate to view such concepts as dominance and its complications (e.g. incomplete penetrance and expressivity) as secondary to the random phenotype concept in *formal* genetic theory, while of course representing biological phenomena of particular genetic interest (cf. KEMPTHORNE 1971).

Historically, the random phenotype concept was introduced clearly for the case of dichotomous phenotypes by YULE (1902, pp. 228–237) with discussion of its implementation in data analysis. This work seems to have been consistently neglected. For the case of quantitative traits (reviewed e.g. in MATHER and JINKS 1971, pp. 3–7), the important results of JOHANNSEN (1903) were of course based just on the classical factor concept, with breeding tests confirming purity of lines. These and other results suggested the *multiple factor hypothesis* which became central in subsequent developments of biometrical methods and theory, particularly in combination with the new concept of the effective factor (as distinct from the classical factor). It appears that these biometrical developments have in effect diverted attention from the possibilities of basing analysis of data which are quantitative (or possibly multivariate, etc.) on the classical factor concept. Such possibilities have only been appreciated since YULE, so far as the present writer is aware, in the present work, largely completed in 1959 but unpublished hitherto; in very incomplete forms, in the data analyses of GREEN

(1954, pp. 615–616) and BRUELL (1962), as discussed in sections 3 and 4 below; for the case of real-valued data, in MODE and GASSER (1972 and references therein), discussed in section 3 below; and in quite general form in COLLINS (1967 and subsequent references cited below).

The preceding comments concern the *formal* applicability of the classical factor concept and corresponding data-analytic methods. In the case of any given genetic material, it is an empirical question, and usually also a partly-theoretical question, whether application of this concept, or the effective factor concept, may be appropriate, interesting, or useful. MATHER and JINKS (1971) survey much of the experience and theory usually cited to support a judgment that it is unlikely that just one or several classical factors might account for the inheritance of some phenotypic character being investigated. (cf. FISHER, IMMER and TEDIN 1932, p. 107.) In any case, since new kinds of possibly heritable characters are frequently considered, it seems important to keep in view the unrestricted feasibility in principle of analysis based on the classical factor concept. (This idea seems implicit in comments of JOHANSEN 1903, pp. 21–22, 1911.) Of course fruitful implementation of such possibilities requires concomitant consideration of choice of specific phenotypes and related questions of experimental design and data interpretation.

In some cases (both new and old) it may be of interest to consider both kinds of analysis. This may be the case when there is interest simultaneously in the somewhat distinct kinds of theoretical and practical goals served by the two methods; or when there is interest in exploring possible relations between specific results of the two methods. The examples below represent classical analyses of results previously published with effective factor analyses, and include some qualified comparisons of results; however these provide at best rudimentary illustrations of the possibilities just referred to. In particular the examples below are *ad hoc* since the experiments were not conceived nor designed with a view to classical analysis. (For comments from several viewpoints on the recognized limitations of theoretical and technical scopes of the effective factor concept and/or the biometrical approach, see MATHER and JINKS 1971; MATHER 1949, especially pp. 38–40, 149; MATHER 1954; ESPINASSE 1942; REEVE and WADDINGTON 1952, especially papers and discussion by MATHER and WOOLF; FALCONER 1960, especially pp. 129–134; KEMPTHORNE 1971.)

Example: The central idea may be illustrated by a simple hypothetical example. Suppose that MENDEL had found no experimental material with distinct constant phenotypes (a condition he emphasized as necessary “to avoid . . . risk of questionable results”), but had been able to work only with phenotypes having distributions over a common range (even in the case of distinct parental varieties). (Actually some of MENDEL’s phenotypes were continuous measurements or graded qualities. He was able to dichotomize these to obtain “constant” qualitative characters only because the ranges of respective distributions did not overlap, particularly in the case of his parental varieties.) Specifically, suppose that Mendel’s homozygote AA gives trait W (wrinkled pea) with probability .7 (and hence gives trait $\bar{W}$, smooth pea, with probability .3). Suppose similarly that

homozygote aa gives W with probability .1, and that heterozygote Aa gives W with probability .5. On Mendel's assumption, any second (or subsequent) filial generation progeny has, with respective probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$, one of the genotypes AA , Aa , aa . Correspondingly any such progeny has trait W with probability $\frac{1}{4}(.7) + \frac{1}{2}(.5) + \frac{1}{4}(.1) = .45$. Similarly, any second backcross progeny has trait W with probability $\frac{1}{2}(.1) + \frac{1}{2}(.5) = .3$. Similarly backcrossing n times to aa gives progeny each having trait W with a probability approaching .1 as n increases.

One can expect fairly close agreement between such probabilities and corresponding observed proportions of W 's, and thus a fairly clear confirmation of Mendel's assumptions if they hold in such a case, with data no more extensive than Mendel's (with only small probabilities of misleadingly large discrepancies due to sampling fluctuations). More conclusive confirmation (comparable with JOHANNSEN's 1903 pure strain breeding-test) is obtainable, for example, by observing progeny of each individual selected from the backcross mentioned, through successive generations of further backcrossing to aa 's. About half of such individuals, in any moderately large sample, will (under the Mendelian assumption) be aa and hence will breed true; among any such an individual's offspring, the observed proportion of W 's may be judged for compatibility with the theoretical probability .1, allowing for sampling fluctuations.

Section 2. Mendelian models for random phenotypes: We adopt a usual notation for probability models (cf. e.g. FELLER 1968):

$$\begin{aligned} x: & \text{any specified phenotype;} \\ S = \{x\}: & \text{the set of all phenotypes among which distinctions} \\ & \text{will be made} \\ R: & \text{any specified subset of } S; \\ P(R): & \text{the probability, under given conditions, of } R. \end{aligned} \quad (2.1)$$

To specify that some given genetic conditions G hold (in addition to some given conditions of development and observation of an individual), we write

$$P(R|G) . \quad (2.2)$$

For example, if P_1 denotes a given true-breeding parental strain, then $P(R|P_1)$ denotes the probability that an individual in that strain will develop with a phenotype in R . It will suffice to discuss here the special but important case of discrete distributions, in which case $P(R|G)$ is determined in the familiar way by the elementary probability distribution function (p.d.f.) $P(x|G)$, $x \in S$.

We shall use the term *phene* to refer to such a probability distribution, and to the biological potential of which it is a model. When for some x , $P(x|G) = 1$, we have the special case of a *constant* (non-random) phene (or constant phenotype). It is frequently convenient, as below, to represent a p.d.f. by a point in some appropriate space, with coordinates which are the probabilities of respective points (phenotypes) x :

$$P[G] = (P(x_1|G), P(x_2|G), \dots) ,$$

where $x_1, x_2, \dots$ denote the points $x \in S$ in some given order. An observational counterpart (estimate) of a phene $P[G]$ will be denoted by $P^*[G]$; for example, $P^*(R|G) = n(R|G)/n(G)$, where $n(G)$ is the number of G individuals observed, and $n(R|G)$ is the number among these having a phenotype in R .

For the simplest Mendelian model, in which a *single factor* is assumed, consider breeding from true-breeding parental strains P_1 and P_2 . The respective genotypes may be denoted as usual: $P_1 = AA$, $P_2 = aa$, and $F_1 = Aa$. Since any B_1 (first backcross, $P_1 \times F_1$) individual is of genotype either P_1 or F_1 , with probabilities each $\frac{1}{2}$, B_1 may be called a *random genotype*. This admits the convenient *symbolic* representation

$$B_1 = \frac{1}{2}P_1 + \frac{1}{2}F_1. \quad (2.3)$$

The corresponding phene is given by the analogous equation $P(R|B_1) = \frac{1}{2}P(R|P_1) + \frac{1}{2}P(R|F_1)$ for each R , or more concisely

$$P[B_1] = \frac{1}{2}P[P_1] + \frac{1}{2}P[F_1]. \quad (2.4)$$

This illustrates the quite general way in which *any* Mendelian model, which is characterized just by the standard (linear) relations (such as (2.3)) at the genotypic level, is represented empirically just by analogous linear restrictions (such as (2.4)) among phenes. Because of the simplicity of this link with standard Mendelian relations, it is not necessary to extend this part of our discussion in general terms. Additional cases are represented schematically in Figure 1, and in the applications below. In the case of one-factor models, these illustrate the general feature that an equation such as (2.4) characterizes each generation bred subsequent to the first filial generation F_1 , as having a phene which is just a certain weighted average of the phenes of the parental and F_1 generations. Analogous features are found in more general Mendelian models including cases of linkage.

When it is of interest to represent each phenotype x (completely or partially) by a corresponding number (a real-valued *measurement*) $y = y(x)$, then it follows immediately from (2.4) that, for each number y ,

$$\text{Prob}(Y \leq y|B_1) = \frac{1}{2}\text{Prob}(Y \leq y|P_1) + \frac{1}{2}\text{Prob}(Y \leq y|F_1), \quad (2.5)$$

and hence that the corresponding expected values (assuming their existence) satisfy

$$E(Y|B_1) = (\frac{1}{2})E(Y|P_1) + (\frac{1}{2})E(Y|F_1). \quad (2.6)$$

(This coincides with the *first* equation in MATHER's 1949, p. 42, standard development of quantitative inheritance theory. In such treatments there do not appear the more general equations (2.4) and (2.5), nor corresponding general theoretical considerations.) Equations (2.5) and (2.6) illustrate the general fact that any linear restriction among probability functions, such as (2.4), is "inherited" by any linear functional of the probability functions.

It is convenient to extend our formal use of the term phene to distributions $P[G]$ in which G is a random genotype. For example,

$$P[B_1] = P[\frac{1}{2}AA + \frac{1}{2}Aa] = \frac{1}{2}P[AA] + \frac{1}{2}P[Aa] .$$

(Of course there is in general no direct biological counterpart of such a phene.)

A schematic example of random-phenotypic expression of the one-factor model is represented in Figure 1. For illustrative simplicity here, just three phenotypes are distinguished, denoted respectively by 1, 2, and 3. Each point ($P(1|G)$, $P(2|G)$) in the figure represents a phene

$$P[G] = (P(1|G), P(2|G), P(3|G)) ;$$

since $P(3|G) = 1 - P(1|G) - P(2|G)$, it is ignored for graphic simplicity.

In other examples the number of phenotypes distinguished will often exceed three, so that in general a phene $P[G]$ is not representable as a point in a space of two (or even three) dimensions. However in all cases, under the *assumption* of a *one-factor* (two-allele) model, each phene $P[G]$ is a point in a triangle with vertices $P[P_1]$, $P[P_2]$, and $P[F_1]$. In this sense Figure 1 serves as a convenient schematic representation of the general random-phenotypic structure of the one-factor (two allele) Mendelian model. We note that no such model can hold unless all the points $P[G]$ are co-planar.

Of course when only constant phenotypic traits are considered, there are at

FIGURE 1.—One factor model with random phenotypes: case of three phenotypes.

most three phenotypes (just two in cases of dominance). In these cases the triangle P_1, P_2, F_1 coincides with the triangle $(0,1), (1,0), (0,0)$; or, in a case of dominance, $P_2 = (0,1)$ and $P_1 = F_1 = (1,0)$, representing a degenerate triangle.

When the model is generalized to allow two or more factors, the dimensionality required for a representation like that of Figure 1 increases rapidly. For *any* Mendelian model, the dimensionality D of the analogue of Figure 1 is at most one less than the number of distinct genotypes represented in a given experiment. Of course D may vary with choice of the set $S = \{x\}$ of phenotypes; the number of phenotypes considered, less one, is another upper bound on D . Due to sampling variation, data points $P^*[G]$ will in general fail to satisfy exactly the linear restrictions which hold among corresponding true points $P[G]$ when a given Mendelian model is assumed.

Interpretations of data: From the standpoint of mathematical statistics each Mendelian model is a statistical hypothesis characterized by some set of equations such as (2.4) or (2.5). In general the sample space has any form, and the respective distributions (phenes) may be subject to no other restrictive assumptions ("nonparametric"). Thus many cases of possible genetic interest correspond to problems of mathematical statistics which have not yet been considered systematically. Even in the special case of real-valued data typified by (2.5), that equation alone represents a generalization or extension, not yet systematically investigated, of the much-studied "two-sample nonparametric" problems. The following sections include illustrations of some ways in which methods available for simpler problems can be adapted for use here.

Questions of compatibility of data with a specified Mendelian model are formulated and treated in several conceptually disparate ways in current research literature: 1. Bayesian (e.g. SMITH 1963, 1968); 2. Likelihood (but non-Bayesian; e.g. EDWARDS 1970, 1971, 1972); more widely established methods: 3. Tests formulated with consideration of power (e.g. MORTON 1962); 4. Tests and estimation-procedures formulated more traditionally (e.g. BAILEY 1961). The methods of data interpretation represented in the following sections fall into categories 3 and 4, but may be better described as relatively unformalized judgments in research contexts, based in part on standard techniques, particularly confidence region estimation. (The writer's general views on such questions are given in BIRNBAUM 1961, 1969, 1970, 1971).

Section 3. Mendelian analysis of quantitative data: When a breeding experiment of the usual (diallel cross) kind provides a (real-valued) measurement for each individual, each phene is represented as in (2.5) above by a c.d.f. $F(y|G)$. Each of these has an empirical counterpart, the empirical c.d.f. (e.c.d.f.) $F^*(y|G)$, for each generation G observed, and it is these which represent the available data.

A standard approach to questions of compatibility of such data with, for example, equation (2.5), is based on the use of a statistic of the Chi-square type to provide a formal test of the statistical hypothesis (2.5). But these methods are subject to well-known questions concerning the partly-arbitrary division of the range of observations into intervals, as well as to more basic questions about

the reasonableness of *any* formal test. We may apply instead estimation methods which lend themselves to interpretations in research contexts. For example, if some division of the range of measurements into intervals is adopted, then instead of computing a formal test significance level, one can examine and compare graphically, for each interval, the "theoretical" and "observed" probabilities (as these are designated in some expositions of the Chi-square method) as each is represented (estimated) by the superimposed "theoretical" and "observed" frequency histograms, with each ordinate accompanied by a graphical indication of its precision. Such graphs allow convenient interpretations of data in genetic contexts, comparable with the interpretations in Section 4 below for cases of multinomial data.

Another kind of statistical approach is based on a direct observational analogue of (2.5) above, for example upon a statistic such as

$$d = \sup_y |d(y)| \quad (3.1)$$

where

$$d(y) = F^*(y|B_1) - \frac{1}{2}F^*(y|P_1) - \frac{1}{2}F^*(y|F_1). \quad (3.2)$$

Since this statistic represents absolute values of discrepancies between "observed" and "theoretical" e.c.d.f.'s, it (or some variant of it) would be a plausible basis for a formal test of the statistical hypothesis (2.5). As mentioned above, such test procedures have not been developed. A partial exception is the ingenious method proposed by MODE and GASSER (1972) which achieves formal validity of a test by requiring special and unusual conditions on the design of a breeding experiment, or else by restricting analysis to a subsample of the available data.

We may apply, instead of any formal test, less formal graphical methods as follows: On the same graph plot the "observed" empirical c.d.f. $F^*(y|B_1)$ (an estimate of the left member of (2.5)) and the confidence band represented by the boundaries

$$F^*(y|B_1) \pm d_{B_1} \quad \text{for all } y; \quad (3.3)$$

and also the "theoretical" e.c.d.f.

$$F^*(y|B_1) = \frac{1}{2} [F^*(y|P_1) + F^*(y|F_1)] \quad (3.4)$$

(which is an estimate of the right member of (2.5)); and also the confidence band represented by the boundaries

$$\frac{1}{2} [F^*(y|P_1) + F^*(y|F_1) \pm (d_{P_1} + d_{F_1})]. \quad (3.5)$$

If the bands overlap for *each* y in the range of possible observations, then some c.d.f. is compatible with both bands, and compatibility of the data with the Mendelian assumption (2.5) is indicated; otherwise incompatibility is indicated. Concerning the strength of such statistical evidence, suppose that the confidence levels associated with the two bands are .95 and .90, respectively. Then simple calculations show that $1 - (.95)(.90) \doteq .15$ is an *upper bound* for the significance level of the indicated testing procedure. For the similar procedure based on the (much wider) bands with levels respectively .99 and .98, we find $1 - (.99)(.98) \doteq .03$ as the upper bound on the significance level. Apart from such formal

testing considerations, graphical consideration of one or more pairs of such bands allows convenient interpretations in context.

The indicated confidence bands for $F(y|B_1)$ are determined directly by use of available tables or approximations (see for example LINDGREN 1960–1962, pp. 300–304, 400). By *considering* the same method, as it *could* be applied also to give separate estimates of $F(y|P_1)$ and $F(y|F_1)$, we determine the constants d_{P_1} and d_{F_1} which appear in (3.5) above. These, and d_{B_1} , each depend on the respective sample sizes n_G and on respective chosen confidence levels, say $1 - \alpha_G$, where G is B_1 , P_1 , or F_1 . The band represented by (3.5) above has, it is easily verified, a confidence level *bounded below* by $(1 - \alpha_{P_1})(1 - \alpha_{F_1})$. The examples above illustrate these relations, and the considerations which may guide one or more specific possible choices of values of α_G and d_G in an application.

BRUELL's data on behavior in mice: Some elements of the random phenotype concept were applied to complement the usual scaling methods in BRUELL's (1962) study of the inheritance of behavior in mice: Under a simple one-factor model, it follows from equation (2.5) above that the range of values observed among B_1 individuals will be just the combination (union) of ranges observed among P_1 and F_1 individuals. BRUELL compared ranges thus deduced with observed ranges, to test the adequacy of such models (p. 61, col. 1; p. 65, col. 1).

At other points (pp. 64, 66), because the mean values required for the usual scaling methods cannot be computed (due to experimental truncation of the distribution of observations), BRUELL adopted the median as an alternative parameter of central tendency; he then used observed medians in place of means in relations like (2.6), to judge adequacy of a one-gene model. Here however a different relationship is appropriate (as consideration of the underlying equation (2.5) readily shows) since the median of measurements on B_1 's is *not* in general equal to the average of the respective medians of measurements on P_1 's and on F_1 's.

A major part of BRUELL's analysis is the application of the biometrical approach to estimate the numbers of effective factors underlying inheritance of certain observed behavior traits in experimental mice.

An application to BRUELL's data of the methods indicated above based on e.c.d.f.'s showed inconclusive results. More precisely, the four experiments each included just B_1 , B_2 , and F_2 generations, for which Mendelian assumptions such as (2.5) could be judged for compatibility with the data. In none of the 12 generations thus considered was the upper bound on significance levels found to be low. This reflects only in part the conservative approximate character of such bounds. Consideration of the graphs of the confidence bands shows that the sample sizes (195 or fewer, in each generation) are too small for .95 bands narrower than $\pm d = \pm .1$ for any *single* c.d.f. It is a matter of behavioral-genetic judgment whether there is possible scientific interest or significance in possible discrepancies (between "observed" and "theoretical" distributions) of the magnitudes compatible with such graphs (or data). In any case, the low precision of the estimates does not allow any conclusive judgments of incompatibility of the data with even the simple one-factor classical Mendelian model.

(For the different, biometrical analysis of BRUELL, some broadly analogous limitations of precision and conclusiveness appear.) Of course questions of desired precision, and conclusiveness with respect to specific kinds of possible cases, should be considered as far as possible among questions of experimental design and research strategy.

Section 4. Mendelian analysis of trinomial and some other multinomial data: For data of the multinomial form, several approaches may be considered. A convenient account of various techniques available for estimation and testing multinomial parameters has been given by MILLER (1966, pp. 215–219), with particular reference to simultaneous confidence interval estimation or testing of several specified linear functions of multinomial parameters. The latter are applicable to the respective parameters by which discrepancies from Mendelian assumptions may be represented.

As an example of such parameters, we may write

$$\Delta(x|B_1) = P(x|B_1) - \frac{1}{2}[P(x|P_1) + P(x|F_1)] \quad (4.1)$$

to denote, for each x , the magnitude of discrepancies between the left and right members of the Mendelian one-factor condition (2.4). For the general case $\Delta(x|G)$ is defined analogously. Thus equations such as (2.4) can be stated in the form $\Delta(x|G) = 0$.

Judgments of compatibility of data with Mendelian assumptions may be guided by confidence interval estimates of such parameters. In many cases each such confidence interval is represented adequately by the estimate $\Delta^*(x|G)$, obtained from (4.1) and its analogues by substituting estimates P^* for the respective probabilities P , accompanied by the estimate of its standard error (calculated in the usual simple way). Such interpretations of data guided by confidence intervals subsume the usual formal test of a hypothesis, whose result is indicated by the inclusion or exclusion of the value zero. Such use of intervals facilitates other equally important interpretations less conveniently coordinated with formal tests, including in particular judgments about plausible magnitudes of systematic errors (including those of the kinds discussed by SMITH 1968, p. 100), and judgments about possible biological importance or interest associated with estimated magnitudes of parameters.

An initial example may be given in terms of estimates based on GREEN's data (Cross 1), introduced more systematically below. The estimates

$$P^*(25|P_1) = .922, P^*(25|F_1) = .011, P^*(25|B_1) = .219$$

determine the estimate $\Delta^*(25|B_1) = 0.25$, and an associated estimated standard error less than .03. The large magnitude of this estimated discrepancy (.25), from a theoretical (one-factor model) estimate of a probability of a certain phenotype ($x = 25$), would presumably be of genetic interest and significance, if the experiment itself had been conceived with a view to classical Mendelian analysis. Systematic errors of this order of magnitude are no doubt implausible. The relatively small estimated standard error (below .03) indicates the implausibility of attributing the estimated discrepancy primarily to sampling variability.

In discussing further examples, we shall use similar but briefer phrases to refer to all of the kinds of considerations mentioned with this example.

Such considerations can be applied easily to each comparison between a direct and a theoretical estimate in the case of binomial data (i.e. the case of dichotomous phenotypes, illustrated briefly by an introductory example in Section 1 above).

In cases where just three phenotypes are distinguished (or in cases where for some purposes only three categories of phenotypes will be distinguished), such considerations and interpretations can be based conveniently on graphs representing direct and theoretical estimates, with their respective precisions indicated by confidence regions. For this case Figure 1 represents a hypothetical example of a one-factor Mendelian model. An observational analogue of such a figure is obtained by plotting estimates of phenes,

$$P^*(i|G) = n(i|G)/n(G), \quad \text{for } i = 1 \text{ and } 2, \quad (4.2)$$

where $n(i|G)$ is the number of individuals with phenotype i , among the $n(G)$ observed individuals of genotype G . We shall call these *direct estimates*. Examples appear in the figures below, where each dot *not* enclosed in a small circle is a direct estimate of a phene.

The precision of such an estimate may be indicated by the magnitude and shape of a confidence region estimate, which may be determined with adequate precision for present purposes by extending to the trinomial case the method of MOSTELLER and TUKEY (1949, pp. 185–186). Such confidence regions are represented by the roughly elliptic contours around direct estimates of phenes in the figures below. Each dashed contour represents a 99% confidence region, and each solid contour a 95% region.

The assumption that a one-factor model holds provides additional *theoretical estimates*. For example by substituting in (2.4) the estimates (4.2) for the parental and first filial phenes we obtain the theoretical estimate

$$\bar{P}^*(i|B_1) = \frac{1}{2}[P^*(i|P_1) + P^*(i|F_1)], \quad i = 1, 2. \quad (4.3)$$

Analogous theoretical estimates are obtained similarly for each generation other than the parental and first filial. Such estimates appear in the figures as dots enclosed by small circles and labeled, for example, B_1 as an abbreviation for

$$(\bar{P}^*(1|B_1), \bar{P}^*(2|B_1)).$$

The precisions of some of the theoretical estimates are represented similarly. It sufficed for the applications here to determine these just in certain cases, where only one term in the right member of an equation such as (4.3) has appreciable variability; this occurs when each remaining term represents a direct estimate very near a vertex of the triangle. Then the contour around the theoretical estimate is a simple projection of the contour around that direct estimate with appreciable variability.

Applications to GREEN's data on skeletal variation in the mouse: E. L. GREEN (1951, 1954, 1962) has reported the results of a series of experimental crosses

TABLE 1
Identification of strains and crosses

Symbol:	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇
Strain:	P	C3H	DBA	C57BL	NB	BALB/s	SEC/2
Cross number	P ₁	P ₂		Reference by E. L. GREEN			
0	S ₄ (C57)	S ₆ (BALB)		1951, p. 398, Table 3			
1	S ₁ (P)	S ₇ (SEC/2)		1954, p. 615, Table 5			
2A	S ₂ (C3H)	S ₄ (C57)		1962, p. 1089, Table 7			
2R	S ₂ (C3H)	S ₄ (C57)		1962, p. 1089, Table 6			
3	S ₁ (P)	S ₅ (NB)		1954, p. 614, Table 3			
4	S ₃ (DBA)	S ₆ (BALB)		1962, p. 1089, Table 5			
5	S ₃ (DBA)	S ₄ (C57)		1962, p. 1088, Table 4			
6	S ₂ (C3H)	S ₆ (BALB)		1962, p. 1089, Table 8			
7	S ₅ (NB)	S ₇ (SEC/2)		1954, p. 614, Table 4			
8	S ₂ (C3H)	S ₃ (DBA)		1962, p. 1088, Table 3			

among seven inbred strains of mice, with interpretations based on the biometrical approach including estimates of numbers of effective factors. His interpretations will be compared below with those obtainable by classical Mendelian analysis. For simplicity of notation and reference we shall identify the respective strains of mice by the symbols S_i defined in Table 1, which also describes the respective experimental crosses and published sources of the data.

Five phenotypes, called skeletal types, are distinguished, after simplifications discussed by GREEN (1954, p. 611). The number of presacral vertebrae is either 25, 26, or 27; or else the form is asymmetrical, being either intermediate between 25 and 26, or intermediate between 26 and 27. For the asymmetrical types we shall use the respective symbols 25.5 and 26.5 in place of GREEN's symbols A_1 and A_2 . Thus the notation for our phenic models will be:

$P(x|G)$ = the probability of skeletal type x , under given
genotypic conditions G (and some given develop-
mental conditions), where $x = 25, 25.5, 26, 26.5,$
or 27 .

For example, $P(26|S_1 \times S_3)$ denotes the probability of 26 presacral vertebrae in an offspring F_1 of a cross between strains S_1 and S_3 .

As empirical counterparts of these probabilities we have

$$P^*(x|G) = n(x|G)/n(G),$$

where $n(G)$ is the number of G individuals observed, and $n(x|G)$ is the number of x individuals among these. For example, the first entry in the table for Cross 1 ($S_1 \times S_7$; GREEN 1954, p. 615, Table 5) gives $P^*(25|S_1) = .922 = n(25|S_1)/n(S_1)$, where $n(S_1) = 384$. It will suffice to illustrate most of our interpretations by graphical representations of some of GREEN's tables.

The initial example discussed above illustrates a feature found throughout the

FIGURE 2.—Cross 2A: Comparison of results with estimates based on one factor model.

data from all crosses: Every estimate $P^*[G]$ satisfies exactly or very nearly one of the conditions

$$\begin{aligned} P(26.5|G) = P(27|G) = 0, \text{ or} \\ P(25|G) = P(25.5|G) = 0. \end{aligned} \quad (4.4)$$

This feature allows convenient graphical approximate representations of all the data, since instead of considering the range of our observed points $P^*[G]$ to be the 5-dimensional unit hypercube, or its 4-dimensional subset determined by the condition that probabilities add to unity, we can take the range of observed points to be (approximately) just two 2-dimensional faces of the 5-dimensional unit hypercube, namely (a) the face whose points have coordinates $(P(25|G), P(26|G))$; and (b) the face whose points have coordinates $(P(26|G), P(27|G))$. For example the data for Cross 3 show that $P^*[P_1]$ lies in the first such face, in which it is represented exactly and completely by the coordinates

$$(P^*(25|P_1), P^*(26|P_1)) = (.922, .034) ,$$

FIGURE 3.—Cross 3: Comparison of results with estimates based on one factor model.

as plotted in Figure 3, where it is labeled simply P_1 .

Any data point $P^*[G]$ not lying in either of these faces is represented by the nearest point (projection) in the nearest face. For example in Cross 3 the last two coordinates of the point $P^*[F_2]$ are very near 0; this point is represented in Figure 3 by a point labeled F_2 , which is its projection into that face. In some cases (such as B_2 in Figure 4) an observed point lying near both faces has been represented by its projections in each face.

When all data points of a cross lie in or very near one face, as occurs except in Crosses 1, 4, and 6, then the graph lies just in that face. In the remaining cases, the graph lies in two adjacent faces, which can be represented as in Figure 4, parts 1 and 2 for Cross 4 (as if we fold open and flatten the surfaces of the hypercube on which data points have been plotted).

Appraisal of one-factor models. Cross 1: This cross is not represented by a figure, but Figure 4 (in two parts) for Cross 4 typifies adequately the features

to be discussed. We note that the estimated P_1 and P_2 points lie in different triangles, with neither near the common edge. Hence the theoretical (one-factor) estimate for F_2 lies well in the interior of the cube:

$$F_2 = \frac{1}{4} P_1 + \frac{1}{2} F_1 + \frac{1}{4} P_2 \doteq (.24, .02, .53, .04, .17).$$

But this appears incompatible with the direct estimate $(.01, .02, .96, .00, .01)$, which lies very near the vertex $(0,0,1,0,0)$ common to the two triangular figures. A conclusion of incompatibility with a one-factor model was also reached in the more detailed consideration of the B_1 data from this cross, in the initial example following (4.1) above.

Cross 2A: An arrow in Figure 2 marks the discrepancy between the direct estimate B_1 and the theoretical estimate $\bar{B}_1$. The magnitude of this discrepancy is small relative to the magnitude of sampling variability associated with B_1 , as indicated by the confidence regions around B_1 . Thus the Mendelian condition (2.4) appears quite compatible with these estimates. This conclusion is only

FIGURE 4.—Cross 4: Comparison of results with estimates based on one factor model.

further confirmed by adding consideration of the variability of the theoretical estimate. Analogous comments apply to the discrepancies associated with B_2 and F_2 . Thus all data here are highly compatible with a one-factor model. Moreover this conclusion does not reflect merely consideration of the magnitudes of sampling variabilities, since the parental and F_1 estimates span a wide range of the phenic space.

Cross 3: Here the respective discrepancies between B_1 and F_2 and corresponding theoretical estimates remain large even after account is taken of the latter's sampling variabilities, not indicated in Figure 3. Hence the one-factor model is not plausible here, unless it is judged that systematic errors may account for discrepancies of such magnitudes. If only much smaller systematic errors are plausible, one may consider whether the discrepancies from the one-factor model may be of a kind and degree having possible genetic interest.

Somewhat formal tests of statistical hypotheses such as (2.4) may be con-

sidered subsumed among the preceding interpretations. For example if (2.4) holds then the probability is $(.99)^2 \doteq .98$ that the true phene $P[B_1]$ is included in both of the 99% regions associated respectively with the direct estimate B_1 and the theoretical estimate $\bar{B}_1$. The probability is somewhat larger that the two regions overlap (whether or not one or both includes the true phene). Hence non-overlapping may be taken as statistical evidence against (2.4), "at a significance level less than .02".

Cross 5: The figure (not reproduced here) and interpretations are intermediate between those for Crosses 2A and 3. Sampling variability alone seems to be a marginally plausible explanation of the observed discrepancies from the one-factor model.

Other crosses: Similar considerations appear in the remaining crosses. Interpretations similar in most respects to those given for Cross 2A also apply to Crosses 7 and 8. Interpretations like those for Cross 3 also apply to Crosses 0, 4, and 6.

Other models: For crosses giving results incompatible with one-factor models, analogous interpretations concerning two-factor models with possible linkage were based on a combination of graphical and more formal methods. These are indicated in summary in Table 2. The effective precision of the data is relatively limited, with respect to the larger number of parameters which characterize two-factor models. Hence only tentative indications of inadequacy of two-factor models were found in some cases.

Plausibility of a scale: When all of the data points observed in the nine crosses are plotted (projected) in a single two-part figure like Figure 4, accompanied by 95% confidence contours as indications of precision, the resulting figure (not reproduced here) strongly suggests that all of the phenes considered may lie on or very near a single smooth curve, connecting by a shallow arc the respective vertices where $P(25) = 1$ and $P(26) = 1$, and then connecting by a similar arc the respective vertices where $P(26) = 1$ and $P(27) = 1$. A specific function (curve) of this kind, which fails to intersect relatively few of the confidence contours, is given by

TABLE 2

Cross	Estimated minimal numbers	
	Effective factors†	Factors
0	3	2*
1	21	2 or 3*
2A	2	1
3	4	2
4	2	2 or 3*
5	2	1 or 2
6	6	2 or 3*
7	1	2
8	0	1

* Subject to possible increase with completion of data analysis.

† Cf. GREEN 1951, pp. 404-405; 1954, p. 622; 1962, pp. 1092-1093.

$$\begin{aligned} (P(25), P(26)) &= (\Phi(t), 1 - \Phi(t + \frac{1}{2})) \\ (P(26), P(27)) &= (\Phi(t), 1 - \Phi(t + \frac{1}{2})) \end{aligned}$$

for $-\infty < t < \infty$, where as usual Φ denotes the standard normal cumulative distribution function. Thus it is plausible that the many (unprojected) phenes P represented in these experiments all lie in a relatively narrow zone including the indicated one-dimensional continuum. This pattern is striking here because the range of mathematically possible phenes P is a four-dimensional continuum. On the other hand, the reader familiar with the literature on quantitative and graded traits may well have regarded the preceding classical Mendelian analysis, in which the natural morphological ordering of the skeletal types played no role, as an exercise in the suspension of belief in the reality of such an ordering, as well as a way of avoiding those arbitrary and hypothetical assumptions which characterize biometrical (scaling and threshold) analyses as distinct from classical Mendelian analysis.

The function indicated above represents a typical case of the inverse normal transformations of GREEN's data which are a basis for his biometrical analysis (GREEN 1962, p. 1094). The preceding analysis gives independent support, derived quite differently from the same data, for the plausibility of existence of an underlying biological continuum; and for the plausibility of existence of a scale on that continuum on which unknown biological effects are approximately additive, independent, and of comparable magnitudes. The latter constitutes the theoretical basis for GREEN's biometrical analyses and estimates of numbers of *effective* factors, summarized in Table 2 (cf. GREEN 1951, pp. 404-405; 1954, p. 622; 1962, pp. 1092-1093). Despite the essential theoretical differences between the effective factor and the classical factor concepts, there are striking qualitative similarities, and even certain quantitative similarities, between respective estimates compared in the table. This is of course plausible on consideration of the conceptual links between the biometrical and classical approaches; but it would seem hazardous to predict such correspondences in new genetic material. Perhaps in some cases where such correspondences are found they may support interpretations of a more deeply integrated kind.

The writer is indebted to Dr. EARL L. GREEN for the opportunity to develop the methods reported here as a Visiting Investigator at the Jackson Laboratory, Bar Harbor, Maine, during the summer of 1959; and for his continuing advice and encouragement. Helpful suggestions concerning earlier versions of this material were also provided by Dr. ROBERT L. COLLINS, members of the Genetics Seminar of Columbia University and the Statistics Seminar of the Johns Hopkins University, and Mr. LADD McLINDEN. Parts of the work were completed with kind assistance by the staff of the Laboratorio Internazionale di Genetica e Biofisica, Naples, and by the Library staff of the Marine Biological Laboratory, Woods Hole. The writer is also indebted to Professor KEI TAKEUCHI for planning the method and calculations of the approximate trinomial confidence regions; and to Mr. McLINDEN, Mr. HERMAN FRIEDMAN of the IBM Systems Research Institute, Mr. J. WANDERLING, and Mr. R. SAMPSON for additional assistance on computational methods.

LITERATURE CITED

BAILEY, N., 1961 *Introduction to the Mathematical Theory of Genetic Linkage*, Oxford.

- BIRNBAUM, A., 1959 Notes on analysis of skeletal cross data. Memorandum of July 10, 1959, to Dr. Earl Green and Mr. Richard Sampson, The Roscoe B. Jackson Memorial Laboratory, Bar Harbor, Maine. Typed, pp. 1-10. —, 1961 Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. *J. Am. Stat. Ass.* **56**: 246-249. —, 1969 Concepts of statistical evidence. In: *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, Edited by S. MORGENBESSER, P. SUPPES and M. WHITE, St. Martin's Press, N.Y. —, 1970 Statistical methods in scientific inference. (Copy containing proof corrections omitted from publication, available from author.) *Nature* **225**: 1033. —, 1971 A perspective for strengthening scholarship in statistics. *Am. Stat.* **16**: 14-17.
- BUELL, J. H., 1962 Dominance and segregation in the inheritance of quantitative behavior in mice. In: *Roots of Behavior*, Edited by E. L. BLISS. Harper, N.Y.
- COLLINS, R. L., 1967 A general nonparametric theory of genetic analysis. I. Application to the classical cross (Abstract) *Genetics* **56**: 551. —, 1968a A general nonparametric theory of genetic analysis. II. Digenic models with linkage for the classical cross. (Abstract) *Genetics* **60**: 169-170. —, 1968b On the inheritance of handedness. I. Laterality in inbred mice. *J. Heredity* **59**: 9-12. —, 1969a On the inheritance of handedness. II. Selection for sinistrality in mice. *J. Heredity* **60**: 117-119. —, 1969b A general nonparametric theory of genetic analysis. III. Definition of a robust realized heritability, h^2_c . (Abstract) *Genetics*, **61**(Supplement): s11-s12. —, 1970a A new genetic locus mapped from behavioral variation in mice: audiogenic seizure prone (*asp*). *Behavior Genet.* **1**: 99-109. —, 1970b The sound of one paw clapping: an inquiry into the origin of left-handedness. In: *Contributions to Behavior-Genetic Analysis—the Mouse as Prototype*, Edited by G. LINDZEY and THIESSEN, Appleton-Century-Crofts. —, 1970c A general nonparametric theory of genetic analysis. Manuscript.
- COLLINS, R. L. and J. L. FULLER, 1968 Audiogenic seizure prone (*asp*): a gene affecting behavior in linkage group VIII of the mouse. *Science* **162**: 1137-1139.
- EDWARDS, A. W. F., 1970 Estimation of the branch points of a branching diffusion process (with discussion). *J. Royal Stat. Soc. Ser. B* **32**: 155-174. —, 1971 Estimation of the inbreeding coefficient from ABO blood-group phenotype frequencies. *Amer. J. Human Genet.* **23**: 97-98. —, 1972 *Likelihood*. Cambridge: Cambridge University Press.
- ESPINASSE, P. G., 1942 The polygene concept. *Nature* **149**: 732.
- FALCONER, D. S., 1960 *Introduction to Quantitative Genetics*. Ronald Press, N.Y.
- FELLER, W., 1968 *An Introduction to Probability Theory and its Applications*, 3rd ed. Wiley, N.Y.
- FISHER, R. A., F. R. IMMER and O. TEDIN, 1932 The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics* **17**: 107-124.
- GREEN, E. L., 1951 The genetics of a difference in skeletal type between two inbred strains of mice. *Genetics* **36**: 391-409. —, 1954 Quantitative genetics of skeletal variation in the mouse. I. Crosses between three short-ear strains (P, NB, SEC/2). *J. Natl. Cancer Inst.* **15**: 609-624. —, 1962 Quantitative genetics of skeletal variation in the mouse. II. Crosses between four inbred strains (C3H, DBA, C57BL, BALB/c). *Genetics* **47**: 1085-1096. —, 1971 The threshold model: uses and limitations. Pp. 143-152. *Proc. of the Conf. on Craniofacial Growth, 1967, Univ. of Mich., Ann Arbor*. Edited by I. N. MOYERS and W. KROGMAN. Pergamon Press, Oxford and New York.
- JOHANNSEN, W., 1903 *Ueber Erblichkeit in Populationen und in Reinen Linien*. Jena: Gustav Fischer. (Partial translation in PETERS 1959.) —, 1911 The genotype conception of heredity. *Am. Naturalist* **45**: 129-159.
- KEMP THORNE, O., 1971 Discussion, p. 152 of GREEN 1971.

- LINDGREN, B. W., 1960–1962 *Statistical Theory*. Macmillan, N.Y.
- MATHER, K., 1949 *Biometrical Genetics*, Dover, N.Y. —, 1951 *The Measurement of Linkage in Heredity*, 2nd edition, Methuen, London. —, 1954 The genetical units of continuous variation. *Proc. 9th Int. Congr. Genet.* 1: 106–123.
- MATHER, K. and J. L. JINKS, 1971 *Biometrical Genetics*, 2nd ed. Cornell University Press, Ithaca.
- MILLER, R. G., 1966 *Simultaneous Statistical Inference*. McGraw-Hill, N.Y.
- MODE, C. J. and D. L. GASSER, 1972 A distribution-free test for major gene differences in quantitative inheritance. *Mathematical Biosciences* 14: 143–150.
- MORTON, N. E., 1962 Segregation and linkage. In: *Methodology in Human Genetics*. Edited by W. J. BURDETTE, Holden-Day.
- MOSTELLER, F. and J. W. TUKEY, 1949 The uses and usefulness of binomial probability paper. *J. Amer. Statist. Assn.* 44: 174–212.
- PETERS, J. A., 1959 *Classic Papers in Genetics*. Prentice-Hall, N.J.
- REEVE, E. C. R. and C. H. WADDINGTON, 1952 *Quantitative Inheritance*. Her Majesty's Stationary Office, London.
- SMITH, C. A. B., 1963 Testing for heterogeneity of recombination fraction values in human genetics. *Ann. Hum. Genet.* 27: 175–182. —, 1968 Testing segregation ratios. Pp. 99–130. In: *Haldane and Modern Biology*, Edited by K. R. DRONAMRAJU. J. Hopkins University Press, Baltimore.
- YULE, G. U., 1902 Mendel's laws and their probable relations to intra-racial heredity. *New Phytol.* 1: 194–207, 222–238.