XXVIII.—Point Binomials and Multinomials in relation to Mendelian Distributions. By John Brownlee, M.D., D.Sc.

(Read November 20, 1911. MS. received February 16, 1912.)

THE subject of this paper is a consideration of the mathematics of some of the problems given by present-day developments of Mendelism. Breeding formulæ are becoming more and more complex, and it seems probable that many others are yet more complex. Without a suitable mathematical analysis it will be nearly impossible to analyse some of these by direct experiment, while a suitable analysis must always render the method of experimental attack much less obscure. Some of the formulæ in the present paper were calculated when I read my last paper * on this subject to this Society, but most of the developments seemed too remote from actual experimental work to render their publication useful. However, a paper published in the last number of the *Journal of Genetics*, by H. M. Leake, concerning the hybridisation of the cotton plant suggests that the time has come when they may prove of practical value.

Before approaching the mathematics, however, I think that some remarks on the notation of Mendelism may be profitably made. This notation seems to me specially cumbrous. When the elements are indicated by letters they are regularly written in sequence. Thus an organism containing three pairs of elements is denoted by (*aa bb cc*). If this mate with another of similar constitution but different properties, (AA BB CC), all the possible combinations of these appear in the second generation, and it becomes very fatiguing to the eye and brain to read the notation. I think that it would be preferable if the different elements which can combine were written in parallel rows. We have then the two arrangements above described denoted by

$$\left|\begin{matrix} aa \\ bb \\ cc \end{matrix}\right| \qquad \left|\begin{matrix} AA \\ BB \\ CC \end{matrix}\right| ,$$

and when we come to consider such combinations as

$$\left|\begin{matrix} aA \\ BB \\ cC \end{matrix}\right| ,$$

* "The Inheritance of Complex Growth Forms on Mendel's Theory," *Proc. Roy. Soc. Edin.*, vol. xxxi. p. 251.

each acting part of the combination can be read at a glance. This will take up a little more space in printing, but the greater clearness will more than compensate.

So much for notation. There is, however, also a very easy use of symbolical multiplication which allows of the numbers of each special combination being immediately written down. As the F_2 generation is a complete picture of all possible combinations, this generation can always be at once written down directly.

If

$$\left|\begin{matrix} aa \\ bb \end{matrix}\right| \quad \text{and} \quad \left|\begin{matrix} \mathrm{AA} \\ \mathrm{BB} \end{matrix}\right|$$

are two races where a changes with A and b with B, we have the stable race, as far as (aa) and (AA) are concerned, represented by

$$(aa) + 2(a\mathrm{A}) + (\mathrm{AA})\,;$$

likewise that of (bb) and (BB), represented by

$$(bb) + 2(b\mathrm{B}) + (\mathrm{BB}).$$

If these be multiplied symbolically together, the numbers of any combination can be at once seen.

Thus those combinations containing the element (aa) are

$$\left|\begin{matrix} aa \\ bb \end{matrix}\right| \quad 2\left|\begin{matrix} aa \\ b\mathrm{B} \end{matrix}\right| \quad \left|\begin{matrix} aa \\ \mathrm{BB} \end{matrix}\right|,$$

and likewise the same holds for the other two terms.

In the same way symbolical multiplication may be applied to the mating of special cases. Thus, take the case given by Bateson on p. 193 of the current number of the *Journal of Genetics*, where the mating of (Ff Pp Ii ff Pp Ii) is discussed. Here, writing each gamete in columns, we have for the first eight possible gametes of the first parent

$$\left|\begin{matrix} \mathrm{F} \\ \mathrm{P}, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} \mathrm{F} \\ \mathrm{P}, \\ i \end{matrix}\right. \left|\begin{matrix} \mathrm{F} \\ \mathrm{P}, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} \mathrm{F} \\ \mathrm{P}, \\ i \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ i \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ i \end{matrix}\right.$$

and for the four possible gametes of the second parent, each duplicate,

$$\left|\begin{matrix} f \\ p, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ \mathrm{I} \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ i \end{matrix}\right. \left|\begin{matrix} f \\ p, \\ i \end{matrix}\right.$$

Multiplying symbolically, we have thirty-two different combinations. The first sixteen will be each unique, but the second half

$$\left\{ \left|\begin{matrix} f \\ p\,, \\ \mathrm{I} \end{matrix}\right. + \left|\begin{matrix} f \\ p\,, \\ \mathrm{I} \end{matrix}\right. + \left|\begin{matrix} f \\ p\,, \\ i \end{matrix}\right. + \left|\begin{matrix} f \\ p\,, \\ i \end{matrix}\right. \right\}^2 .$$

With a little practice all the special forms existing may be easily worked out.

The general theorems on which this paper is based are given in the following lemmas. These in the form in which they are shown, so far as I know, are new, but they are so simple that it seems unlikely they have not been proved many times before. They depend on the well-known formulæ for the transference of the moments of a body round the centre of gravity to a point at a distance h therefrom, and on the values of the moments of a point binomial round the centre of gravity. If μ_2, μ_3, μ_4 be the moments of a curve round the centre of gravity, and μ'_2, μ'_3, μ'_4 be the moments round a point at a distance h, then

$$\begin{aligned} \mu'_2 &= \mu_2 + h^2 , \\ \mu'_3 &= \mu_3 + 3h\mu_2 + h^3 , \\ \mu'_4 &= \mu_4 + 4h\mu_3 + 6h^2\mu_2 + h^4 . \end{aligned} \qquad \text{(A)}$$

Also the moments of $(q+1)^n$ about the centre of gravity are given by

$$\begin{aligned} \mu_2 &= \frac{c^2nq}{(q+1)^2} , \\ \mu_3 &= \frac{c^3nq(q-1)}{(q+1)^3} , \\ \mu_4 &= \frac{c^4nq}{(q+1)^2} \left\{ 1 + \frac{3(n-2)q}{(q+1)^2} \right\} . \end{aligned}$$

Lemma I.—If an expression of distributed terms be multiplied by a second expression of distributed terms, the moments of the compound expression round the centroid vertical are given by

$$\begin{aligned} \mu_2 &= \zeta'_2 + \zeta''_2 , \\ \mu_3 &= \zeta'_3 + \zeta''_3 , \\ \mu_4 &= \zeta'_4 + \zeta''_4 + 6\zeta'_2\zeta''_2 , \end{aligned}$$

where ζ'_2, ζ''_2, etc., are the moments of the separate expressions, and μ_2, etc., those of the compound expression. Let the second expression be denoted by $a+b+c+ \ldots$; then the centre of gravity of the compound expression is moved a distance equivalent to the distance of the middle point of the new expression from its centre of gravity. For the first can be conceived as concentrated at its centre of gravity, and if its mass be M, the distribution of the new expression will be equivalent to $M(a+b+c+\ldots)$. We have therefore to obtain the moments of the compound expression round the new centre of gravity.

Let h_1, h_2, h_3, etc., be the distance of a, b, c, etc., from the centre of gravity then by the preceding formulæ (A)

$$M(a+b+c+\ \ldots)\mu_2 = M\zeta^2\Sigma a + M\Sigma ah^2\,,$$

or

$$\mu_2 = \zeta'_2 + \zeta''_2\,;$$

likewise

$$M(a+b+c+\ \ldots)\mu_3 = M\zeta_3\Sigma a + M\zeta_2\Sigma ah + M\Sigma ah^3 = \zeta'_3 + \zeta''_3\,,$$

since $\Sigma ah = 0$ by definition. So also

$$\mu_4 = \zeta'_4 + \zeta''_4 + 6\zeta'_2\zeta''_2\,.$$

And in general

$$\mu_2 = \Sigma\zeta'_2\,,$$
$$\mu_3 = \Sigma\zeta'_3\,,$$
$$\mu_4 = \Sigma\zeta_4 + 6\Sigma\zeta'_2\zeta''_2\,.$$

Lemma II.—The even moments μ_2 and μ_4 of $(1+q)^m\ (q+1)^n\ (1+1)^q$ $(1+p)^k\ (p+1)^l$ etc., are constant if $m+n=$ const. and $k+l=$ const., and the odd moment μ_3 depends on the difference $m-n$, $k-l$, etc. This follows at once from the preceding since the second moment of $q+1$ is the same as those of $1+q$, while the third moment of these is equal but of opposite sign.

Lemma III.—If $(1+q)^m\ (q+1)^n\ (1+1)^p\ (1+p)^r\ (p+1)^s$ etc., be a distribution, then the first moments round an horizontal axis are the same if $m+n=c$, $n+s=c'$, etc. For, consider the distribution given by $a+b+c+\ldots$, where a, b, and c are posited at equal distances m.

The first moment is

$$\frac{\dfrac{a^2}{m^2}+\dfrac{b^2}{m^2}+\dfrac{c^2}{m^2}+}{a\ +\ b\ +\ c\ +\ \ldots}\,.$$

Multiplying the distribution by $1+q$ and $q+1$ respectively, and limiting to three terms which is sufficient, we have the two expressions

$$a+(b+qa)+(c+qb)+qc,$$
$$qa+(qb+a)+(qc+b)+c.$$

The first moments are then

$$\frac{a^2+(b+qa)^2+(c+qb)^2+q^2c^2}{m^2(a+b+c)}\,,$$

and

$$\frac{q^2a^2+(qb+a)+(qc+b)^2+c^2}{m^2(a+b+c)}\,,$$

which are obviously identical.

The forms which the different point binomials tend to assume are interesting. As already shown by Professor Pearson, $(q+1)^n$ has the same slope as the curve given by his Type III., namely, by

$$y=\left(1+\frac{x}{a}\right)^p e^{-\gamma x}, \quad \text{where} \quad \gamma=\frac{p}{a}.$$

This curve has for the criterion

$$2\beta_2-3\beta_1-6=0,$$

where

$$\beta_1=\frac{\mu_3^{\,2}}{\mu_2^{\,3}} \quad \text{and} \quad \beta_2=\frac{\mu_4}{\mu_2^{\,2}}.$$

When the criterion is greater than zero, Type IV. arises. As just shown, the second and fourth moments of $(q+1)^p$ are the same as those of $(q+1)^m (1+q)^n$, when $m+n=p$; so that if the criterion be zero for the curve which represents the limit of $(q+1)^p$, it will be greater than zero for the curve representing $(q+1)^m (1+q)^n$, since the third moment of the latter is less than that of the former in the ratio $m-n$ to $m+n$, or, in other words, a curve having the same moment relationships as Type IV. will represent the result of mixed dominance.

I have not been able to prove that the limit of $(q+1)^m (1+q)^n$ is represented by

$$y=\frac{y_0}{\left(1+\frac{x^2}{a^2}\right)^m} e^{-\nu \tan^{-1}\frac{x}{a}},$$

but I think that it is highly probable that it is so, and, in addition, even if the limit is different, the curve must be nearly the same as that given by the Type IV. equation. This at once renders futile the class of criticism which, ignorant of mathematical principles, condemns Type IV. as inapplicable to practical problems on account of the imaginary roots in the denominator.

The form $(1+1)^n$ leads to the normal curve, as is well known.

The form $(1+n+1)^p$ has the moment relationships of Type IV., as shown in my former paper, and finally $(a+1+a)^p$ those of Type II.

The foregoing lemmas apply directly to the groupings which are possible on the Mendelian hypothesis. Four possibilities or combinations of possibilities have till now been experimentally ascertained or premised as the result of experiment:—

1. Blending;
2. Dominance;
3. Partial dominance;

and 4. Coupling or duplication of parts.

The first of these leads to groupings represented by the binomial or multinomial $(1+1)^p$ or $(1+0+1)^p$, etc. The second leads to groupings represented by $(3+1)^p$ or $(3+0+1)^p$, etc. The third to groupings represented by such multinomials as $(1+0+2+0+0+1)^p$, where the number of zeros is unequal on the two sides of the middle significant term. The fourth are represented by the groupings $1+2(n-1)+1$ and $\overline{n-1}+2+\overline{n-1}$, either alone or in combination, when n is a positive power of 2.

1. *Blending.*—Blending is known in many instances, as when a red cow mating with a white gives a roan; but of quantitative blending, to which the present analysis applies, I have not found many examples. One is given by H. M. Leake in the paper already referred to.* In this case he shows that when plants with a different leaf factor are mated the leaf factor of the cross is the arithmetical mean. When his diagram is examined it is found that he mated plants of a certain range of leaf factor with others of a like range of higher value, with a distance between the two approximately equal to the range of each type. For a first approximation each type can be represented by the distribution $1+2+1$, so that the F_2 generation will be given by

$$(1+2+1+0+0+1+2+1)^2,$$

or

$$1+4+6+4+1+2+8+12+8+2+1+4+6+4+1.$$

This is approximately what is obtained on experiment, but there is more compression, so that, although in fact the units of each type fell inside the limits chosen, many probably contained elements which, when pure, ranged beyond these limits. If we write, instead of the above,

$$(1+2+1+0+1+2+1)^2,$$

or

$$1+4+6+4+3+8+12+8+3+4+6+4+1,$$

the result will be nearly that shown by the diagram.

2. *Dominance.*—If dominance exist, the proportions in which the numbers will turn up is as 3 to 1, so that the general formula for a mixed dominance will be $(3+1)^m\ (1+3)^n$. This in general approximates to Type IV. Data on which to make analysis are very scarce, so scarce that I have not thought it worth while to carry out the arithmetical work necessary to test individual cases. There is, however, one case given by Mr Leake in the *Journal of Genetics* (*loc. cit.*), concerning the length of the vegetative period in hybrid cottons, which is sufficiently interesting to require an approximate solution. Here dominance obviously comes specially

* *Journal of Genetics*, vol. i. p. 227.

from one side. If the distribution given by Mr Leake be divided into six portions containing each four of the units and fitted to $(3+1)^3(1+3)^2$, we have the following comparison :—

Actual	+5	12	55	168	95	31	15	1·5
Theoretical	0	10	69	152	109	32	4	

This is a very good fit, considering the nature of the experiments, for the data do not admit of any very accurate analysis. The length of the vegetative period is dependent on the season, and the original crossings have obviously been selected, so that any better result could hardly be hoped for.

It is to be noted in this connection that the standard deviation of $(3+1)^m(1+3)^n$ is constant if $m+n=$ a constant, so that it affords a strict means of comparison between symmetrical and skew distributions if dominance holds. The degree of mixture of race cannot, however, be compared by this means, if blending holds in one case and dominance in another. The simplest case is that of the races blending with regard to four qualities, and is represented by $(1+1)^4$. The same when mixed dominance holds is $(3+1)(1+3)$; the extreme range is the same in both instances by hypothesis, yet in the former case $\mu_2=1$, and in the latter $\mu_2=1·5$, so that between two equally mixed races a 50 per cent. difference might be observed.

Concerning cases 3 and 4 there are no data, so that further discussion is at present unnecessary.

Some remarks may be made in conclusion regarding the methods which should be observed in making experiments. Though no race is approximately pure, it is most probable that the extremes of each race are more pure than the members between. In making cross-fertilisation experiments, then, the extremes should be chosen, as in that way the mixtures will be much more easily analysed mathematically. When, as in Mr Leake's experiments, individuals are chosen from each race by special selection, it is hardly likely that results will be obtained which give much hope of satisfactory analysis.

[*Table*

Table of the Moments of the Principal Forms discussed.

(These, for any complex series, are sufficient by help of the preceding formulæ.)

Form.	μ^2.	μ^3.	μ^4.
$(1+1)^n$	$\dfrac{c^2n}{4}$	0	$\dfrac{c^4n}{4}\left\{1+\dfrac{3(n-2)}{4}\right\}$
$(q+1)^n$	$\dfrac{c^2nq}{(q+1)^2}$	$\dfrac{c^3nq(q-1)}{(q+1)^3}$	$\dfrac{c^4n}{(q+1)^2}\left\{1+\dfrac{\overline{3n-2q}}{(q+1)^2}\right\}$
$(1+p+1)^n$	$\dfrac{2nc^2}{p+2}$	0	$\dfrac{2pn+4n(3n-2)}{(p+2)^2}c^4$.

Appendix on Special Cases of Blending.

If a race with two elements denoted by (aa) be crossed with a race (bb) we have the stable population given by the proportions

$$(aa)+2(ab)+(bb).$$

This gives twice as many organisms having the mean quality as either extreme. In like manner take another race of greater stature, say $(cc)+2(cd)+(dd)$, and let them mate at random with the first population. The stable population will obviously consist of the proportions of

$$\begin{aligned}(aa)+(bb)+(cc)+(dd)&\\ +2(ab)+2(ac)+2(ad)&\\ +2(bc)+2(bd)&\\ +2(cd)&\,.\end{aligned}$$

There is a great number of special cases, all, however, agreeing in that the range is between (aa) and (dd), taking these as the extreme types, and that each mixed element has properties equivalent to the arithmetical mean of the elements.

(1) In the first case, let $(bb)=(cc)$ in the quality examined. The grouping will then be as follows:—

$$\begin{aligned}&(aa)+2(ab)+(bb)\\ &\qquad\qquad (cc)+2(cd)+(dd)\\ &\qquad 2(ac)+2(ad)+2(bd)\\ &\qquad\qquad 2(bc)\,,\end{aligned}$$

or

$$1\ +\ 4\ +\ 6\ +\ 4\ +\ 1.$$

In this case the permanent population is approximately normal, and only one mode appears. If the quality depend on two elements in each, if it is defined by

$$\left|\begin{matrix}aa\\ bb\end{matrix}\right| \quad \left|\begin{matrix}\mathrm{AA}\\ \mathrm{BB}\end{matrix}\right|,\ \text{etc.},$$

the ordinary theorem of chance gives the distribution

$$(1+4+6+4+1)^2,$$

and so on.

(2) In this case, let $(ab)=(cc)$. The grouping is then as follows:—

$$\begin{array}{ccccccc} (aa) & & +2(ab) & & +(bb) & & \\ & & (cc) & & +2(cd) & & +(dd) \\ & 2(ac) & & 2(bc) & & 2(bd) & \\ & & & 2(ad) & & & \end{array}$$

or

$$1 + 2 + 3 + 4 + 3 + 2 + 1,$$

or

$$(1+1+1+1)^2.$$

Again a unimodal curve shows itself. If the quality depends on n pairs of elements, then the general distribution is given by the multinomial,

$$(1+2+3+4+3+2+1)^n.$$

This quickly approximates to the normal curve.

(3) Let (cc) fall between (aa) and (ab); the grouping is then:—

$$\begin{array}{cccccccc} (aa) & & & +2(ab) & & +(bb) & & \\ & & (cc) & & +2(cd) & & & +(dd) \\ & 2(ac) & & & 2(cb) & & 2(bd) & \\ & & & & 2(ad) & & & \end{array}$$

or

$$1 + 2 + 1 + 0 + 2 + 4 + 2+0 + 1 + 2 + 1,$$

or

$$(1+1+0+0+1+1)^2,$$

or we have in this case a multimodal curve.

(4) Let (cc) fall between (ab) and (bb); the grouping is then:—

$$\begin{array}{cccccccccc} (aa) & \ldots & 2(ab) & \ldots & (bb) & & & & & \\ & & & & (cc) & \ldots & 2(cd) & \ldots & (dd) & \\ & 2(ac) & & \ldots & 2(bc) & \ldots & 2(bd) & & & \\ & & & & 2(ad), & & & & & \end{array}$$

$$1 + 0+0 + 2 + 2 + 0 + 1 + 4+1+0+2 + 2+0+0+1.$$

or

$$1+0+0+1+1+0+0+1)^2,$$

which again ultimately approaches normality. This curve, composed of two nearly equal races, has three modes.

(5) Let (cc) be greater than (bb), and we have the following, when a dot indicates a gap of one unit:—

Case (a).

$$\begin{array}{c}(aa)+2(ab)+(bb)+\quad\cdot\quad+(cc)+2(cd)+(dd),\end{array}$$

or

$$\begin{array}{c}2(ac)+2(bc)+2(bd)\\ +2(ad),\end{array}$$

or

$$1+2+3+4+3+2+1,$$

or Case (2) from a different reason.

Case (b).

$$\begin{array}{ccccccc}(aa)+2(ab)+(bb)+ & \cdot & + & \cdot & + & \cdot & +(cc)+2(cd)+(dd)\\ & 2(ac) & & 2(ad) & & 2(bd) & \\ & & & 2(bc), & & & \end{array}$$

or

$$1+2+1+2+4+2+1+2+1,$$

or

$$(1+1+0+1+1)^2.$$

Case (c).

$$\begin{array}{ccccccccccc}(aa)+2(ab)+(bb)+ & \cdot & + & \cdot & + & \cdot & + & \cdot & + & \cdot & (cc)+2(cd)+(dd)\\ & & & 2(ac) & & 2(ad) & & 2(bd) & & & \\ & & & & & 2(bc) & & & & & \end{array}$$

or

$$1+2+1+0+2+4+2+0+1+2+1,$$

or Case (3) again.

(Issued separately October 31, 1912.)