A THEORY OF THE ASSOCIATION OF CHROMOSOMES IN KARYOTYPES, ILLUSTRATED BY DR. PATRICIA JACOBS' DATA

J. O. IRWIN
STATISTICAL RESEARCH UNIT
MEDICAL RESEARCH COUNCIL, LONDON

1. Introduction

Geneticists are interested in studying karyotypes. A karyotype is a microphotograph of a human cell just prior to mitotic division, taken under certain standard conditions. They are able to distinguish between the main groups of chromosomes in the Denver classification [2], and in particular they are interested in a phenomenon called "association" (sometimes it is called "overlap") when several chromosomes approach particularly close together. Figure 1 shows two instances of association. These associations are concerned with members of groups 13–15 and groups 21–22 of the Denver scale. All the data used in this paper refer to these two groups. In what follows, groups 13–15 are called group D and groups 21–22 are called group A. Dr. Patricia Jacobs examined for association a number of karyotypes (each for a different person) from (i) normal individuals (ii) mongols and (iii) abnormals excluding mongols. Nearly all the members of group (iii) had Klinefelter's syndrome, but three were mosaics and one was a case of Turner's syndrome. These people are abnormal in sexual development, Klinefelter's are all of male types; cases of Turner's syndrome are of female type; one mosaic was of male type and two of female types. Groups (i) and (ii) both have four A's and six D's; but the mongols have five A's and six D's that is, an extra chromosome. This is supposed to be the cause of their characteristic abnormalities.

2. Basic model for associations of chromosomes in karyotypes

Now let us consider, say, five chromosomes in the field of observation. They may all five be separate; two may be associated and three separate; two may be associated and the other three associated, and so on. Clearly, each possible case corresponds to a partition of the number 5. The question examined here is whether the associations may be accounted for by the hypothesis that the

Adviser in Biometric Techniques (retired).

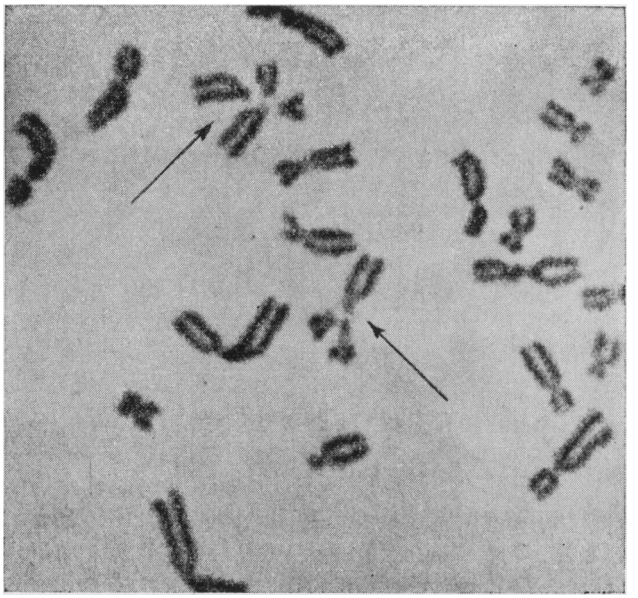

FIGURE 1

Karyotype illustrating chromosome associations.
The upper part of the figure shows four chromosomes in association;
two members of group 13–15 and two members of group 21–22,
while the lower shows an association of three chromosomes,
one member of group 13–15 and two members of group 21–22.

chromosomes are moving about randomly in the field of observation with a fixed probability p, say, of any two being associated.

In general, suppose that there are n chromosomes moving randomly in the field of observation, that the probability of any pair being observed to be associated is p, and that this is the same for all pairs. Suppose further that the probability of any single chromosome being associated with any one group of already associated chromosomes is also p. This is obviously a very much simplified model. It assumes that a group of chromosomes is regarded as associated if its members all fall within a small area a of fixed size and that the probability of a further chromosome falling in this area is also p, and that a is independent of the number in the group. However, the model enables us to calculate the probability distribution of all the types of association that can occur (including no association as one type) and seems to lend to a distribution which closely describes the observed facts.

In each karyotype examined, every association was classified by its partitional type. The numbers in each group are given in table I.

Any particular set of associations in a single cell can of course be represented by a partition of n. An example is helpful. Suppose there are five chromosomes.

TABLE I

NUMBER OF ASSOCIATIONS OF KARYOTYPES BY GROUP

Klinefelters (12) are all of male type; cases of Turner's syndrome (1) of female type; one mosaic was of a male type, the other two of female type.

Group	No. of Karyotypes	No. of Associations Male	Female	Total
(i) Normals (4 A, 6 D)	29	615	277	822
(ii) Mongols (5 A, 6 D)	19	234	302	536
(iii) Abnormals excluding Mongols (4 A, 6 D)	16			726

The partitions of 5 are (1^5), (21^3), (31^2), (41), (32) and (5). The term (1^5) means that there are five separate chromosomes and no associations; the term (21^3) means there is one association of two chromosomes and three singles; the term (31^2) means one association of three and two singles, and so on. The term (5) means that all five chromosomes overlap. The association can clearly be satisfied by the number of parts in the partition.

We use the notation

$$(1-p)_{(3)} = (1-p)(1-2p)(1-3p), \qquad (1-p)_{(2)} = (1-p)(1-2p), \tag{2.1}$$

and so forth. Then it may be shown that the probabilities of the possible partitions are as given in table II. To illustrate the manner of calculating these probabilities, we consider the calculation of the probability for (31^2). Five

TABLE II

PROBABILITIES OF THE POSSIBLE PARTITIONS

Parts	Probability	
$5(1^5)$	$(1-p)_{(4)}$	
$4(21^3)$	$10p(1-p)_{(3)}$	
$3(31^2)$	$10p^2(1-p)_{(2)}$	$25p^2(1-p)_{(2)}$
(2^21)	$15p^2(1-p)_{(2)}$	
$2(32)$	$10p^3(1-p)$	$15p^3(1-p)$
(41)	$5p^3(1-p)$	
$1(5)$	p^4	
Total	1	

different chromosomes can be partitioned into (31^2) in $5!/(3!2!) = 10$ ways. The probability that two particular chromosomes are both associated with a given chromosome is p^2. The probability that a third is not associated with the group of three is $(1-p)$ and that the fourth is not associated with either of the two previous groups (of two and one) is $(1-2p)$. Hence, the probability for

(31^2) is $10p^2(1-p)(1-2p)$. It should be noticed that when the partitions are classified according to the number of parts which they contain, the numerical coefficient in the probability of r parts is $\Delta^r(0^5)/r!$; for example, for three parts $\Delta^3(0^5)/3! = 25$. This provides a check on the coefficients. If N associations are observed the expected numbers are N times the above probabilities. The mean is $\{1-(1-p)^5\}/p$ and p may be estimated by equating this expression to the observed mean $\bar{x}$. This results in an estimate of p with a standard error of

$$\left\{\frac{\text{Var}\,(\bar{x})}{(1+2q+3q^2+4q^3)^2}\right\}^{1/2}, \tag{2.2}$$

where $q = (1-p)$ and Var $(\bar{x})$ is calculated in the ordinary way.

2.1. *Numerical example. Mongols (both sexes) A chromosomes only.* For this case the observed frequencies, the probabilities and expected frequencies, are given in table III.

TABLE III

NUMERICAL EXAMPLE

Observed frequencies, probabilities, and expected frequencies for A chromosomes in Mongols (both sexes).

Parts	Obs. Freq.	Probability	Exp. Freq.
$5(1^5)$	407	$(1-p)_{(4)} = 0.74636$	400
$4(21^3)$	112	$10p(1-p)_{(3)} = 0.23534$	126
$3(31^2)$	10	$10p^2(1-p)_{(2)} = 0.00719$	4
(2^21)	7	$15p^2(1-p)_{(2)} = 0.01079$	6
$2(32)$	0	$10p^3(1-p) = 0.00021$	0.1
(41)	0	$5p^3(1-p) = 0.000105$	0.06
$1(5)$	0	$p^4 = 0.0000006$	0
	536		536

The fit is clearly not unsatisfactory. We have

$$\begin{aligned} p &= 0.0370 \text{ with S.E. } 0.0030 \text{ or } 8\%, \\ \chi^2_1 &= 7.68, \quad P = 0.006. \end{aligned} \tag{2.3}$$

However, of this 7.68, the contribution from the small group with three parts is 4.9.

Reasonably satisfactory fits are obtained in all cases as long as the A and D associations are considered separately. When all associations are considered together, however, that is, for all ten or eleven chromosomes taken together, there is not a satisfactory fit, though there is still a general resemblance between the observed and fitted distributions (see table VII). For example, the excessively improbable associations do not occur at all.

This is partly because the probabilities of the A associations and the D associations differ somewhat (though not significantly in these data in any one case)

and partly because the cross (AD) associations have a different probability from either. I shall endeavor, in a later paper, to find the bipartitional distribution of all associations whether like associations or cross associations.

3. Generalization of the theory

The above theory can be generalized. Corresponding to any partition of u,

$$(\lambda_1^{\pi_1}\lambda_2^{\pi_2} \cdots \lambda_k^{\pi_k}) \text{ with } \sum_{i=1}^{k} \pi_i = \pi, \tag{3.1}$$

the number of parts, the corresponding probability is

$$\frac{n!}{(\lambda_1!)^{\pi_1}(\lambda_1!)^{\pi_2} \cdots (\lambda_k!)^{\pi_k}\pi_1!\pi_2! \cdots \pi_k!} f(p), \tag{3.2}$$

where $f(p) = p^{n-\pi}(1-p)_{(\pi-1)}$ while the probability of all partitions with π parts is $[\Delta^{\pi}(0^n)/\pi!]f(p)$. For example, with $n = 10$ and the partition (32^21^3), the probability is

$$\frac{10!}{3!(2!)^2(1!)^3 2!3!}[p^4(1-p)_{(5)}] = 12600\, p^4(1-p)_{(5)}. \tag{3.3}$$

The other cases for $n = 10$ are easily worked out and column 1 in table VII shows the results for the 892 normal karyotypes all taken together. Only 12 of the 42 possible partitions, namely those with six parts or more were observed; the remaining 30 with five parts or less were not observed, and their collective expectation is only about one tenth of a unit.

While the general resemblance between the observed and expected frequencies is clear, the fit is not satisfactory. There is an excess of cells with no associations; the probability p was estimated from the equation $\bar{x} = (1 - q^{10})/p$ to be 0.0187 with a standard error of 0.0007. When the associations in groups (ii) and (iii) are all taken together, similar results are obtained.

When the A and D associations were considered separately, the fit of observed and expected frequencies was found to be satisfactory, as tables IV–VII show.

TABLE IVa

OBSERVED AND EXPECTED FREQUENCIES OF PARTITIONAL DISTRIBUTIONS FOR NORMAL A

	Males		Females		Both Sexes	
Parts	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
4 (1^4)	536	535.4	228	226.9	764	761.3
3 (21^2)	76	78.5	46	48.2	122	126.6
2 (31)	2	1.2	2	1.1	4	2.3
(2^2)	1	0.9	1	0.8	2	1.7
1 (4)	0	0.007	0	0.009	0	0.2
Total	615	615	277	277	892	892

TABLE IVb

NORMAL D

Parts	Males		Females		Both Sexes	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
6 (1^6)	484	476.4	221	220.8	705	696.5
5 (214)	115	128.5	52	52.6	167	181.7
4 (31^3)	8	3.0	1	1.1	9	4.1
(2^21^2)	6	6.8	3	2.5	9	9.3
3 (41^2)	2	0.04	0	0.01	2	0.05
(321)	0	0.16	0	0.05	0	0.21
(2^3)	0	0.04	0	0.01	0	0.05
2 (51)	0	0.000	0	0.000	0	0.000
(42)	0	0.001	0	0.000	0	0.001
(3^2)	0	0.000	0	0.000	0	0.001
1 (6)	0		0		0	
Total	615	615	277	277	892	892

Table VIII gives the values of $10{,}000p$ and their standard errors. In table VIIIa there are no significant differences between the sexes, either in the normals or the mongols. In the D group but not in the A the value for the mongols is significantly greater than for the normals. The value for the "abnormals excluding mongols" is significantly greater than for the normal males. It is also significantly greater than the average for both sexes in normal as well as in mongol A and in normal D.

In table VIIIb, the values of p can be regarded only as tentative averages for all associations, since the fit of the pooled distributions is unsatisfactory. They show a greater value for the mongols as against the normals, which, it seems, must be mainly due to the D, and a greater value for the abnormals excluding mongols, as against the normals, agreeing with table VIIIa in this respect.

TABLE Va

OBSERVED AND EXPECTED FREQUENCIES OF PARTITIONAL DISTRIBUTIONS FOR MONGOLS A

Parts	Males		Females		Both Sexes	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
5 (1^5)	170	168.4	237	231.6	407	400.0
4 (21^3)	57	60.3	55	65.8	112	126.1
3 (31^2)	4	2.1	6	1.8	10	3.9
(2^21)	3	3.1	4	2.7	7	5.8
2 (32)	0	0.07	0	0.05	0	0.1
(41)	0	0.04	0	0.02	0	0.06
1 (5)	0	0	0	0	0	0
Total	234	234	302	302	536	536

TABLE Vb

MONGOLS D

Parts	Males		Females		Both Sexes	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
6 (1^6)	150	149.9	206	201.1	356	351.0
5 (21^4)	73	73.5	80	89.3	153	162.9
4 (31^3)	5	3.1	5	3.4	10	6.5
(2^21^2)	6	7.0	10	7.7	16	14.7
3 (41^2)	0	0.1	1	0.1	1	0.1
(321)	0	0.3	0	0.3	0	0.6
(2^3)	0	0.1	0	0.1	0	0.1
2 (51)	0	0.001	0	0.001	0	0.002
(42)	0	0.004	0	0.002	0	0.004
(3^2)	0	0.002	0	0.001	0	0.003
1 (6)	0		0		0	
Total	234	234	302	302	536	536

TABLE VIa

OBSERVED AND EXPECTED FREQUENCIES OF PARTITIONAL DISTRIBUTIONS FOR ABNORMALS EXCLUDING MONGOLS

A

Parts	Obs.	Exp.
4 (1^4)	582	575.6
3 (21^2)	132	143.7
2 (31)	9	3.8
(2^2)	2	2.9
1 (4)	1	0.04
Total	726	726

TABLE VIb

D

Parts	Obs.	Exp.
6 (1^6)	523	508.7
5 (21^4)	171	195.4
4 (31^3)	18	6.5
(2^21^2)	10	14.6
3 (41^2)	2	0.1
(321)	0	0.5
(2^3)	1	0.1
2 (51)	0	0.001
(42)	1	0.003
(3^2)	0	0.002
1 (6)	0	
Total	726	726

TABLE VII

OBSERVED AND EXPECTED FREQUENCIES OF PARTITIONAL DISTRIBUTIONS FOR COMBINED GROUPS

All Normals Combined			All Mongols Combined			All Abnormals Combined Excluding Mongols		
Parts	Obs.	Exp.	Parts	Obs.	Exp.	Parts	Obs.	Exp.
10 (1^{10})	418	364	11 (1^{11})	185	126.2	10 (1^{10})	265	207
9 (21^{8})	277	368	10 (21^{9})	152	219.0	9 (21^{8})	239	309
8 (31^{7})	43	22	9 (31^{8})	35	20.1	8 (31^{7})	41	26
($2^{2}1^{6}$)	115	113	($2^{2}1^{7}$)	81	121	($2^{2}1^{6}$)	105	139
7 (41^{6})	5	1	8 (41^{7})	4	1.2	7 (41^{6})	10	1
(321^{5})	20	10	(321^{6})	24	16.7	(321^{5})	30	17
($2^{3}1^{4}$)	8	12	($2^{3}1^{5}$)	31	25.0	($2^{3}1^{4}$)	16	22
6 (51^{5})	1	0.0	7 (51^{6})	1	0.04	6 (51^{5})	4	0.1
(421^{4})	1	0.3	(421^{5})	6	0.7	(421^{4})	3	0.6
($3^{2}1^{4}$)	1	0.2	($3^{2}1^{5}$)	1	0.5	($3^{2}1^{4}$)	2	0.4
($32^{2}1^{3}$)	3	1.0	($2^{4}1^{3}$)	7	1.8	($32^{2}1^{3}$)	8	2.6
($2^{4}1^{2}$)	0	0.4	($32^{2}1^{4}$)	7	3.6	($2^{4}1^{2}$)	1	1.0
5 and less	0	0.1	6 (521^{4})	1	0.53	5 (521^{3})	1	0.3
			($42^{2}1^{3}$)	1		($42^{2}1^{2}$)	1	
			5 and less	0	0	4 and less	0	0
Total	892	892		536	536.4		726	726

Differences between the third group and mongols show up in table VIIIa and are obscured in table VIIIb.

TABLE VIIIa

VALUES OF 10,000p AND STANDARD ERRORS. RESULTS BY CHROMOSOME TYPE AND SEX

	A			D		
Chromosome group:	Males	Females	Both Sexes	Males	Females	Both Sexes
Normals	225 ± 25	320 ± 44	256 ± 22	165 ± 14	147 ± 19	158 ± 11
Mongols	313 ± 36	255 ± 31	280 ± 27	281 ± 27	258 ± 24	268 ± 18
Abnormals excluding mongols	—	—	370 ± 30	—	—	227 ± 15

TABLE VIIIb

VALUES OF 10,000p AND STANDARD ERRORS BOTH TYPES OF CHROMOSOMES TOGETHER AND ALL ASSOCIATIONS INCLUDING AD ASSOCIATIONS

Normals	187 ± 7
Mongols	240 ± 11
Abnormals, excluding mongols	255 ± 10

4. General remarks

Some general remarks may be made.

(1) The distribution obtained in this study is similar in type to the well known "occupied class" distribution. If an object is equally likely to belong to any one of k types, and a random collection of n objects is made, the probability that r types are represented is

(4.1)

$$P(r) = \frac{k!}{r!(k-r)!}\Delta^r(0^n)/k^n = \left(1-\frac{1}{k}\right)\left(1-\frac{2}{k}\right)\cdots\left(1-\frac{r-1}{k}\right)\Delta^r(0^n)/k^{n-r}r!.$$

In the distribution of this study, we can identify p with $1/k$. It is, however, a little more general since in the present case $1/p$ need not be an integer. Biological applications of the "occupied class" distribution have been made before, for instance in W. L. Stevens' study of twinning [3].

(2) In a paper read to this Symposium by Barton, David, Fix, and Merrington [1], use was made of the notion of average distance between chromosomes. It appears that this method and that used here would confirm one another. For example, the authors [1] found a greater average distance between normal than between abnormal chromosomes. This corresponds to a smaller probability of association or overlap, as has been found here.

REFERENCES

[1] D. E. Barton, F. N. David, E. Fix, and M. Merrington, "A review of analysis of karyographs of the human cell in mitosis," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 4, pp. 349–366.

[2] S. Maynard-Smith, L. S. Penrose, and C. A. B. Smith, *Mathematical Tables for Research Workers in Human Genetics*, London, Churchill, 1961, p. 43, figure 1.

[3] W. L. Stevens, "Significance of grouping and a test for uniovular twins in mice," *Ann. Eugen.*, Vol. 8 (1937), pp. 57–73.