

GENETIC CORRELATIONS IN EQUILIBRIUM POPULATION

By RANAJIT CHAKRABORTY

Indian Statistical Institute

SUMMARY. Parent-offspring correlations, correlations between full sibs and half sibs are derived in this paper for populations which is under equilibrium under an arbitrary but fixed mating systems. The different cases with one pair of genes at a locus (autosomal as well as sex-linked) are studied in detail and indication is given for generalisation in the case with multiple alleles.

1. INTRODUCTION

The knowledge of absolute frequencies of the various genotypic combinations of any two relatives with respect to one pair of genes, autosomal or sex-linked, is of great practical value in certain types of studies in human heredity. Various methods of obtaining these frequencies directly or indirectly are prevailing in the literature. The difficulties of the rather straight forward procedure can be well apprehended from Hogben's paper (1933). Li and Sacks (1954) gave a procedure of finding the frequencies of various genotype combinations of near relatives by using the matrices of conditional probabilities. The main purpose of their work was to express such matrices of conditional probabilities of the relatives in the form of linear functions of some basic matrices.

In case one is more interested in the correlations between the relatives rather than the frequencies of different genotype/phenotype combinations, the method of path coefficients, developed by Wright (1921 and later) pays dividend since it gives the correlations instantly once the relationship is specified. This method does not provide the absolute frequencies of the different genotype (or phenotype) combinations.

All of these methods are well known to study the genetic correlations between relatives when the population is assumed to exercise random mating only. Relatively fewer results are known for populations which keep themselves under equilibrium through a more general mating structure.

The object of this study is to derive the genetic correlations between some near relatives in population which is under equilibrium in Wright's sense. We shall study the different cases with one pair of genes at a locus in detail and indicate the the necessary generalizations in the case with multiple alleles at a locus.

2. AUTOSOMAL GENES

2.1 Codominant genes. Genes resting on the locus of our concern may be present in either of the two states, namely A and a . Consider any two individuals and look for the states of the genes.

Definition 1 : Two genes chosen one from each of two unrelated individuals are said to be *alike in state* if they are found to be in the same state.

Definition 2 : Two genes are said to be *identical by descent* if they are replica of the same gene possessed by some ancestor.

From these two definitions it follows immediately that genes of latter kind are necessarily alike in state, excepting the possibility of mutation, but the reverse is not true in general.

Now, a pair of relatives may have both, one, or no genes *identical* by descent depending upon the type of relationship. In case there is exactly one gene common through identity by descent the conditional probabilities that one should be of a certain genotype when the other's genotype is given can be represented by the matrix

$$T_s = \begin{vmatrix} p \left\{ 1 + \frac{2Fq}{p+Fq} \right\} & q \left\{ 1 - \frac{2Fp}{q+Fp} \right\} & 0 \\ \frac{1}{2} \left\{ p + \frac{F(1-2p)}{1-F} \right\} & \frac{1}{2} & \frac{1}{2} \left\{ q + \frac{F(1-2q)}{1-F} \right\} \\ 0 & p \left\{ 1 - \frac{2Fq}{q+Fp} \right\} & q \left\{ 1 + \frac{2Fp}{q+Fp} \right\} \end{vmatrix}$$

for a general mating structure where $F(0 < F < 1)$ is the constant measuring the deviation from randomness of the mating system. Note that this also represents the matrix of transition probabilities of a parent-offspring pair since a parent and one of its offspring always share one gene through identity by descent.

Hence, multiplying the first row of T_s by $p^2 + Fpq$, second row by $2pq(1-F)$ and the third row by $q^2 + Fpq$ one can convert T_s into absolute frequencies for different genotypic combinations of a parent-offspring pair from which the parent offspring correlation (r_{σ_s}) is obtained as

$$r_{\sigma_s} = \frac{1+3F}{2(1+F)} \quad \dots (2.1)$$

(Chakraborty, 1970; also cited in Rao, 1970).

Since we are, for the moment, concerned with autosomal genes only, this will be the correlation for all parent-offspring combinations (mother-son, mother-daughter, father-son or father-daughter). With this correlation alone we can now obtain the correlation between two full sibs or two half sibs through Figs. 1 and 2.

GENETIC CORRELATIONS IN EQUILIBRIUM POPULATION

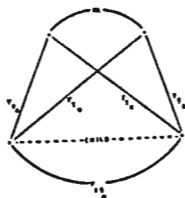


Fig. 1

The correlation between two full sibs

$$r_{FS} = 2r_p^2(1+m) \quad \dots (2.2)$$

 where m , the correlation between the mating partners is given by

$$m = \frac{2F}{1+F}$$

 Inserting this value of m in the above expression (2.2) one gets

$$r_{FS} = \frac{1}{2} \left(\frac{1+3F}{1+F} \right)^2 \quad \dots (2.3)$$

 Putting $F=0$ in (2.3) we get correlations between the full sibs in a random mating population as

$$r_{FS} = 2 \cdot \left(\frac{1}{2} \right)^2 = \frac{1}{2}$$

In Fig. 2 we denote λ to be the correlation between the parents who mate with the common parent. (In the fig., λ is the correlation between the two wives; the husband being the common parent of the two half sibs).

The correlation between the half sibs turns out to be

$$r_{HS} = r_p^2 [(1+m)^2 + \lambda(1+2m)] \quad \dots (2.4)$$

(Rao, 1970).

 Correlations between other relatives also can be worked out similarly using this basic correlation r_p .

2.2. *Genes with dominance relationship.* Let us now assume that the allele A to be dominant over a . Thus, we have with us phenotypes A (consisting of genotypes of AA and Aa) and a (consisting of genotype aa).

The parent-offspring correlation takes the form

$$r_T^{(d)} = \frac{q}{1+q-Fq} + \frac{Fq(1+p-Fp)}{(1+q-Fq)(q+Fp)} \quad \dots (2.5)$$

Needless to say that for $F = 0$, (panmixia) this expression reduces to $\frac{q}{1+q}$ as obtained by Li (1955).

The full sib correlation is easily seen to be

$$r_{FS_s}^{(d)} = \frac{1}{4} + \frac{1}{2} r_{T_s}^{(d)} = \frac{1}{4} + \frac{1}{2} \left[\frac{q}{1+q-Fq} + \frac{Fq(1+p-Fp)}{(1+q-Fq)(q+Fp)} \right] \quad \dots (2.6)$$

It may be recalled that in the absence of any dominance relations between the alleles, the genetic correlations depend only on F , the coefficient of departure from random mating. The correlations are independent of gene frequencies. But in this case the correlations do depend upon the gene frequencies as well. Under panmixia the correlation between the full sibs is given by

$$r_{FS}^{(d)} = \frac{1}{4} + \frac{1}{2} \left(\frac{q}{1+q} \right) \quad [\text{putting } F=0 \text{ in } (2.6)].$$

3. SEX-LINKED GENES

3.1. *Codominant genes.* As usual we taken the homogametic type XX as females and the heterogametic type XY (or XO) as males, where X denotes the sex chromosome. In the usual notations it may be easily verified that the population

$$\begin{pmatrix} A & a \\ p & q \end{pmatrix} \sigma^r \text{ and } \begin{pmatrix} AA & Aa & aa \\ p^2+Fpq & 2pq(1-F) & q^2+Fpq \end{pmatrix} \varphi$$

is in equilibrium under the general mating system. The equilibrium conditions involving the mating frequencies can be derived as follows.

Let the mating matrix be represented by Table 1. We have

$$P_1 = p, P_0 = q$$

and $Q_2 = p^2 + Fpq, Q_1 = 2pq(1-F)$ and $Q_0 = q^2 + Fpq. \quad \dots (3.1)$

TABLE 1. MATING FREQUENCIES WITH SEX-LINKED GENES

females	males		total
	A	a	
AA	u_{11}	u_{20}	Q_2
Aa	u_{11}	u_{20}	Q_1
aa	u_{01}	u_{00}	Q_0
total	P_1	P_0	1

GENETIC CORRELATIONS IN EQUILIBRIUM POPULATION

It is easy to see now that the zygotic proportions of the females in the next generation are given by

$$Q'_2 = u_{21} + \frac{u_{11}}{2}$$

$$Q'_1 = u_{20} + \frac{u_{11}}{2} + \frac{u_{10}}{2} + u_{01}$$

$$Q'_0 = \frac{u_{10}}{2} + u_{00}$$

which in turn give the equilibrium conditions as

$$\text{and } \left. \begin{aligned} u_{11} &= 2u_{20} \\ u_{20} &= 2u_{01} \end{aligned} \right\} \dots (3.2)$$

Note that these conditions also imply that the male frequencies are kept constant in the next generation.

It may be recalled here that for autosomal genes the correlations between the four parent-child pairs are the same, all being $\frac{1+3F}{2(1+F)}$. For sex-linked genes, however, the asymmetrical chromosomal complement of males and females makes it necessary to distinguish the sexes of the relatives. There are four kinds of parent-offspring relationships: father-son, father-daughter, mother-son and mother-daughter. Likewise, there are three kinds of sib-pairs: two-brothers, two-sisters, and brother-sister.

The parent-offspring correlations are easily obtained once we consider the segregation ratios as shown in Table 2. The correlations are as follows:

Father-son correlation, $r_{fs} = 0$.

Father-daughter correlation, $r_{fd} = \sqrt{\frac{1}{2}(1+F)}$

Mother-son correlation, $r_{ms} = \sqrt{\frac{1}{2}(1+F)}$

... (3.3)

and Mother-daughter correlation, $r_{md} = \frac{1+3F}{2(1+F)}$

TABLE 2. MATING TYPES AND SEGREGATION RATIOS WITH SEX-LINKED GENES (After Li, 1961)

parent mother x father	probability of offspring given sex				
	♀			♂	
	AA	Aa	aa	A	a
AA x A	1	0	0	1	0
AA x a	0	1	0	1	0
Aa x A	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
Aa x a	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
aa x A	0	1	0	0	1
aa x a	0	0	1	0	1

From (3.3) we get interestingly enough that there is no correlation between father and son for sex-linked genes, since the son receives his father's *Y*-chromosome. The correlation for mother-daughter is the same as in the case of autosomal genes, because the daughter also receives *p.A* and *q.a* from her father. The two correlations for father-daughter and mother-son are the same. Note that each of these statements are true also in case of random mating populations.

Once these basic correlations are obtained the full sib correlations are obtained from the Fig. 3 as follows :

$$\text{Brother-brother correlation, } R_1 = \frac{1+F}{2}$$

$$\text{Brother-sister correlation, } R_2 = \frac{1+5F+2F^2}{2\sqrt{2}(1+F)} \quad \dots (3.4)$$

$$\text{Sister-sister correlation, } R_3 = \frac{1+F}{2} + \frac{(1+3F)(1+7F+4F^2)}{4(1+F)^2}$$

In deriving this we made use of the fact that correlation between the mating partners is given by

$$m = \sqrt{\frac{2}{1+F}} \cdot F$$

in this case.

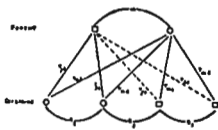


Fig. 3

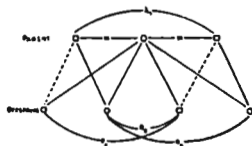


Fig. 4

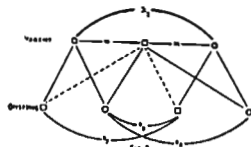


Fig. 5

GENETIC CORRELATIONS IN EQUILIBRIUM POPULATION

From Figs. 4 and 5 one can now establish the following relations for half-sib correlations.

Brother-brother (related through mother) correlation

$$R_4 = \frac{1+F}{2}$$

Brother-sister (related through mother) correlation

$$R_5 = \sqrt{\frac{1+F}{2}} \left[\lambda_1 F + \frac{1+5F+2F^2}{2(1+F)} \right]$$

Sister-sister (related through mother) correlation

$$R_6 = \left[\frac{1+5F+2F^2}{2(1+F)} \right]^2 + \frac{\lambda_1(1+4F+7F^2)}{2(1+F)}$$

Brother-brother (related through father) correlation,

$$R_7 = F^2 + \frac{\lambda_2(1+F)}{2}$$

Brother-sister (related through father) correlation,

$$R_8 = \sqrt{\frac{1+F}{2}} \left[F(1+\lambda_2) + \frac{1+3F}{1+F} \left(\frac{F^2}{1+F} + \frac{\lambda_2}{2} \right) \right] \quad \dots \quad (3.6)$$

and sister-sister (related through father) correlation,

$$R_9 = \frac{1+F}{2} + \frac{F(1+\lambda_2)(1+3F)}{1+F} + \frac{(\lambda_2 + \lambda_2 F + 2F^2)(1+3F)^2}{4(1+F)^2}$$

where λ_1 and λ_2 are the correlations between the two fathers (Fig. 4) and two mothers (Fig. 5) respectively.

3.2. *Genes with dominant relationship.* The four different parent-offspring correlation in this case are obtained as :

$$r_{md} = 1 - \frac{(q+F-3Fq)}{(1+q-Fq)(q+Fp)}$$

$$r_{ms} = r_{fd} = \sqrt{\frac{q+Fq}{1+q-Fq}} \quad \dots \quad (3.7)$$

and

$$r_{fs} = 0.$$

The observations that we made in Section 2.2 are true in this case also.

4. MULTIPLE ALLELES

The foregoing analysis can now readily be generalized to multiple alleles, autosomal or sex-linked. To illustrate the validity of the method we shall consider only the three allelic case. A_1A_1 , A_1A_2 , A_1A_3 , A_2A_1 , A_2A_2 and A_2A_3 represent the different genotypes if the genes are autosomal. For sex-linked genes also these are the genotypes of the homogametic females. In such a case the males will be type A_1 , A_2 or A_3 . Let the gene frequencies be a_1 , a_2 and a_3 respectively where $a_1+a_2+a_3 = 1$.

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS : SERIES B

Before proceeding to derive the correlations it is worth studying the equilibrium conditions in such a situation. The mating frequencies in three allelic cases (autosomal and sex-linked) are represented by Table 3 and Table 4 from which we can easily establish the following necessary and sufficient conditions for equilibrium (in Wright's sense) :

TABLE 3. MATING FREQUENCIES IN 3 AUTOSOMAL ALLELIC CASES

mating	A_1A_1	A_1A_2	A_2A_1	A_2A_2	A_2A_3	A_3A_2	total
A_1A_1	u_{11}	u_{12}	u_{21}	u_{14}	u_{15}	u_{16}	U_1
A_1A_2	u_{12}	u_{22}	u_{22}	u_{24}	u_{25}	u_{26}	U_2
A_1A_3	u_{13}	u_{23}	u_{31}	u_{34}	u_{35}	u_{36}	U_3
A_2A_2	u_{14}	u_{24}	u_{34}	u_{44}	u_{45}	u_{46}	U_4
A_2A_3	u_{15}	u_{25}	u_{35}	u_{45}	u_{55}	u_{56}	U_5
A_3A_2	u_{16}	u_{26}	u_{36}	u_{46}	u_{56}	u_{66}	U_6

TABLE 4. MATING FREQUENCIES IN 3 SEX-LINKED ALLELIC CASES

female	male			total
	A_1	A_2	A_3	
A_1A_1	v_{11}	v_{12}	v_{13}	U_1
A_1A_2	v_{21}	v_{22}	v_{23}	U_2
A_1A_3	v_{31}	v_{32}	v_{33}	U_3
A_2A_2	v_{41}	v_{42}	v_{43}	U_4
A_2A_3	v_{51}	v_{52}	v_{53}	U_5
A_3A_2	v_{61}	v_{62}	v_{63}	U_6
total	V_1	V_2	V_3	1

For autosomal genes :

$$\left. \begin{aligned} u_{22} &= 4u_{14}, & u_{23} &= 2u_{15} \\ u_{33} &= 4u_{44}, & u_{34} &= 2u_{24} \\ u_{35} &= 4u_{16}, & u_{36} &= 2u_{26} \end{aligned} \right\} \dots (4.1)$$

For sex-linked genes :

$$\left. \begin{aligned} v_{21} &= 2v_{12}, & v_{22} &= 2v_{41} \\ v_{31} &= 2v_{13}, & v_{32} &= 2v_{42} \\ v_{33} &= 2v_{41}, & v_{41} &= v_{22} = v_{32} \\ v_{52} &= 2v_{42} \end{aligned} \right\} \dots (4.2)$$

For deriving the correlations it is not enough to follow the weighting scheme of the two allelic case since this superficial approach creates some difficulties as mentioned in Stanton (1960).

The genetic weighting scheme to be followed is that of Stanton (1960) where he places the pure genotypes symmetrically at the vertices of a regular simplex and

GENETIC CORRELATIONS IN EQUILIBRIUM POPULATION

then uses vectorial weights. For three allelic cases the weighting scheme is shown in Fig. 6, where the vectorial weights v_i 's follow the relations

$$v_1^2 = 1, v_1 v_2 = -\frac{1}{2}, \frac{1}{2} v_1(v_1 + v_2) = -\frac{1}{4},$$

$$\frac{1}{2} v_1(v_1 + v_2) = \frac{1}{2}, [\frac{1}{2}(v_1 + v_2)]^2 = \frac{1}{2} \quad \dots (4.3)$$

and $\frac{1}{2}(v_1 + v_2) \cdot \frac{1}{2}(v_1 + v_2) = -\frac{1}{2}$

(Stanton, 1960).

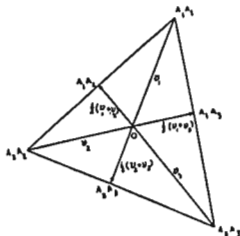


Fig. 6. Full SIB and half SIB Correlations (Autosomal Gene)

It can be seen that the parent-offspring correlations in this case turn out to be the same as those in two allelic case. It is worth noting that Stanton (1960) also observed this for a random mating population.

5. NUMERICAL ILLUSTRATIONS

We shall give as illustrations parent-offspring correlation for characters controlled by one pair of genes with dominance and an hypothetical case with a three allelic autosomal character. For both of them Boorman's (1950) on human blood factors are sufficient.

Though the genetics of the Rh-factor in human blood has advanced greatly in the last two decades for our purpose here we may, however, still treat it as though it were controlled by one pair of genes with dominance. Boorman (1950) reported the data regarding mother-child combinations as shown in Table 5 here.

TABLE 5. MOTHER-CHILD COMBINATION OF Rh BLOOD FACTORS (After L, 1958)

mothers	Rh type of child		total
	(+)	(-)	
(+)	1475	183	1667
(-)	204	129	333
total	1679	311	1990

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS : SERIES B

The parent-offspring correlation, computed from this table, is 0.2853. The theoretical formula (2.6) also leads to the same amount of correlation with an F value of 1.8 percent.

To illustrate the three allelic case we consider Table 6 which is modified from Boorman. Using the same example Stanton (1960) obtained a correlation of magnitude 0.5180 which leads to an estimate of $F = 1.86$ percent.

TABLE 6. PARENT-CHILD ARRAY FOR A-B-O BLOOD TYPES
(After Stanton, 1960)

parent	child						total
	AA	AB	BB	BO	OO	AO	
AA	49	5	—	—	—	122	176
AB	10	10	2	21	—	24	67
BB	—	2	—	4	—	—	6
BO	—	20	4	50	55	29	159
OO	—	—	—	43	622	227	892
AO	122	14	—	28	223	303	690
total	181	181	6	146	901	703	1990

ACKNOWLEDGEMENT

I express my deep gratitude to Professor C. R. Rao who looked after the work with great interest. Thanks are due to Mr. D. C. Rao for his valuable suggestions.

REFERENCES

- BOORMAN, K. E. (1950): An analysis of the blood types and clinical condition of 2000 consecutive mothers and their infants. *Ann. Eugen.*, 15, 120-134.
- CHAKRABORTY, R. (1970): Parent-offspring correlation in an equilibrium population. *Amer. J. Hum. Genet.*, 22, 605-610.
- HOBBEN, L. (1933): A matrix notation for Mendelian populations. *Proc. Roy. Soc. Edin.*, 55, 7-25.
- LI, C. C. and SACER, L. (1954): The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics*, 10, 347-360.
- LI, C. C. (1965): *Population Genetics*, University of Chicago Press.
- (1961): *Human Genetics: Principles and Methods*, McGraw-Hill, London.
- RAO, D. C. (1970): A note on Li's paper. *J. Ind. Soc. Agri. Stat.*, 22, 53-58.
- STANTON, R. O. (1960): Genetic correlations with multiple alleles. *Biometrics*, 16, 235-244.
- WRIGHT, S. (1921): The biometric relations between parent and offspring. *Genetics*, 6: 111-123.

Paper received: April, 1970.