J. Math. Biology (1981) 13: 227–240

Journal of
Mathematical
Biology

Analysis of "Nontraditional" Relationships Under Assortative Mating*

A. Gimelfarb**

Department of Statistics, North Carolina State University, Raleigh, NC 27650, USA

Abstract. The set of conditions on the genetical and developmental mechanisms of quantitative characters as well as on selection and mating system presented in (Gimelfarb, 1981) is expanded, thus enabling one to obtain the genotypic covariances between relatives for a larger variety of relationships. It is also demonstrated that the frequency of a relationship in a population under assortative mating may in general be different from the frequency of this relationship in the population under random mating. A subpopulation of relatives is not necessarily a representative sample of the whole population with respect to the quantitative character distribution. However, for any relationship which is a combination of descendant-ancestor, full sib, Type 1 and Nth uncle-niece relationships, its frequency in a population under assortative mating is the same as in the population under random mating, and the subpopulation of such relatives is a representative sample of the whole population.

Key words: Assortative mating – Covariance between relatives

When one looks at the types of relationships that have been dealt with in the literature in connection with the problem of covariances between relatives, it is interesting to notice that their number is quite limited. Almost all authors, even those who were interested in the problem from a purely theoretical point of view, worked with the list of relationships which comes almost unchanged from Fisher's (1918) original work and includes such relationships as descendant-ancestor, full sibs, half sibs, first cousins and uncle-niece. An attempt to analyze a relationship outside of those "traditional" ones belongs to Fisher himself, who produced an expression for the double first cousin correlation, although without a rigorous proof of it (Fisher, 1918).

A set of conditions on hereditary and developmental mechanisms of quantitative characters as well as on selection and mating system has been established in

* Paper No. 6620 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, North Carolina. This investigation was supported in part by NIH Research Grant No. GM 11546 from the National Institute of General Medical Sciences

** *Present address*: Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA

0303–6812/81/0013/0227/$02.80

the previous paper (Gimelfarb, 1981). Under these conditions, the covariances between any "traditional" relatives can be expressed in terms of quantitative character variance and marital correlations. They are not sufficient, however, for covariances between any relatives (double first cousins is one example) to be expressed in the same terms. In this paper, the previous conditions are supplemented with several additional ones, thus making a new set of conditions that enables the expression of covariances between relatives in terms of the variance and marital correlation for a larger variety of relationships. Some problems pertaining to relationships of different types following from this set of conditions are also discussed.

It will be assumed that the population in every generation goes through two phases: juvenile individuals and reproducing individuals assorted into mating pairs. A quantitative character of an individual is considered to be a sum of two components: hereditary (genotypic value) and nonhereditary (environmental component). Small letters will be used for genotypic values and capital letters will denote phenotypes. The following notations will be used for some distributions.

$P_k^0(Z)$	the distribution of the character among the juvenile individuals in generation k. Its variance is V_k^0.
$p_k^0(z)$	the same as the previous, but for the genotypic values. Its variance is v_k^0.
$p_k(x)$	the distribution of the genotypic values among the reproducing individuals in generation k. Its variance is v_k.
$P_k(X_i, X_j\|(i*j))$	the joint distribution of the characters in mating pairs in generation k. $(i*j)$ stands for the event: individuals i and j mate.
$p_k(x_i, x_j\|(i*j))$	the same as the previous but for the genotypic values.
$F_k^0(Z\|z)$	the distribution of the character in the generation k among the juvenile individuals with the genotypic values z.
$H_k(z\|x_i x_j)$	the distribution of the genotypic values among the offspring of the individuals with the genotypic values x_i and x_j.

Assuming that the genotypic value is a continuous variable, it is easy to see that

$$p_k(x_i) = \int_{x_j} p_k(x_i, x_j|(i*j))\, dx_j,$$

and that the distribution of the genotypic values among the juvenile individuals in the next generation is expressed as

$$p_{k+1}(z) = \int_{x_i} \int_{x_j} p_k(x_i, x_j|(i*j)) H_k(z|x_i, x_j)\, dx_i\, dx_j. \tag{1}$$

All other distributions appearing in this paper will be denoted as Pr[ ]. It will also be assumed for simplicity without loss of generality that the mean genotypic value among either juvenile or reproducing individuals in the population is zero in any generation.

The following is a set of conditions on hereditary and developmental mechanisms, mating system and selection that has been introduced and discussed in the previous paper (Gimelfarb, 1981).

Condition A. The mean genotypic value among the offspring of a mating pair is determined completely by the parental genotypic values as

$$E(z|g_i, g_j) = E(z|x_i, x_j) = \tfrac{1}{2}(x_i + x_j), \tag{2}$$

where g_i and g_j are the parental genotypes.

Condition B. For juvenile individuals, the regressions of the individual's phenotype on the genotypic value and of the individual's genotypic value on the phenotype are linear:

$$E(Z|z) = z, \tag{3a}$$

$$E(z|Z) = h_k^2 Z, \tag{3b}$$

where h_k^2 is the heritability:

$$\frac{v_k^0}{V_k^0} = h_k^2. \tag{3c}$$

Condition C. For reproducing individuals, the regressions of the individual's phenotype on the genotypic value and of the individual's genotypic value on the phenotype are linear:

$$E(X|x) = a_k x, \tag{4a}$$

$$E(x|X) = b_k X, \tag{4b}$$

with

$$\frac{b_k}{a_k} = \frac{v_k}{V_k}. \tag{4c}$$

(It will be assumed further for simplicity that $a_k = 1$ and $b_k = h_k^2$.)

Condition D. The environmental component of a juvenile individual is independent of the parental genotypic values:

$$\Pr[Z|z, x_i, x_j] = F_k^0(Z|z). \tag{5}$$

Condition E. For any two reproducing individuals i and j,

$$\Pr[(i*j)|X_i, X_j, \Omega] = \Pr[(i*j)|X_i, X_j], \tag{6}$$

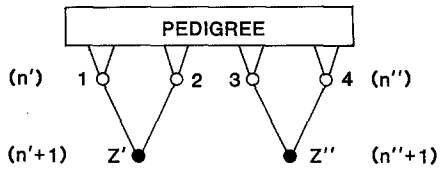
where Ω stands for any set of conditions.

Condition F. The regression of the phenotype of a mating individual on the phenotype of its mate is linear:

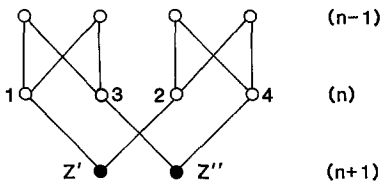
$$E(X_i|X_j, (i*j)) = \rho_k X_j, \tag{7}$$

where ρ_k is the "marital correlation".

The set of Conditions A – F turned out to be sufficient for the covariances between relatives whose relationship is a combination of descendant-ancestor, full sib, Type 1 and Nth uncle-niece relationships to be expressed in terms of the quantitative character variance and marital correlation (Gimelfarb, 1981). These



1



2

Fig. 1. A relationship of general type. (Generation numbers are indicated in parentheses)

Fig. 2. Double first cousins

conditions are not sufficient, however, for the genotypic covariances of other relatives to be expressed in the same terms.

Any relationship other than descendant-ancestor may be depicted as in Fig. 1, where Z' is the phenotype of one of the relatives and Z'' is the phenotype of the other. Individuals 1 and 2 are the parents of one of the relatives, and individuals 3 and 4 are the parents of the other. n' is the generation of the parental pair 1 and 2, and n'' is that of the parental pair 3, 4. Pedigree refers to the pedigree of individuals 1, 2, 3, 4 and includes all connections (either "blood" or "step") existing between these individuals, except for the connections due to matings of individuals 1 and 2, and of individuals 3 and 4. Thus, in the case of "double step sibs" (Fig. 3), Pedigree refers to the matings $(1 * 3)$ and $(2 * 4)$ but not to the matings $(1 * 2)$ and $(3 * 4)$.

Let us now formulate three conditions in addition to A – F.

Condition G. The distribution of the genotypic values among mating pairs, i.e., $p_k(x_i, x_j|(i * j))$ is a bivariate normal.

This is not a very restrictive condition, since many quantitative characters in biological populations are known to have close to normal distributions.

Condition H. The probability for individuals 1, 2, 3, 4, given their genotypic values, to come from a particular Pedigree is unaffected by the matings $(1 * 2)$ and $(3 * 4)$:

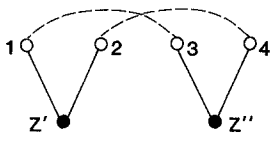
$$\Pr[\mathrm{PED}|x_1, x_2, x_3, x_4, (1 * 2), (3 * 4)] = \Pr[\mathrm{PED}|x_1, x_2, x_3, x_4]. \tag{8}$$

In the case of double-first cousins (Fig. 2), for example, where individuals 1 and 3 as well as 2 and 4 are sibs, this will imply that the environmental components of sibs are independent. Not for every relationship, however, this condition may be readily assumed. It can never hold under assortative mating for a relationship where some of the individuals 1, 2, 3, 4 are ancestors of the others, or for double step sibs (Fig. 3), in which case

$$\Pr[\mathrm{PED}|x_1, x_2, x_3, x_4] = \Pr[(1 * 3), (2 * 4)|x_1, x_2, x_3, x_4]$$
$$\neq \Pr[(1 * 3), (2 * 4)|x_1, x_2, x_3, x_4, (1 * 2), (3 * 4)].$$

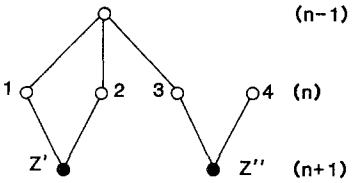
Condition I. The distribution $\Pr[x_1, x_2, x_3, x_4|\mathrm{PED}]$ is a tetravariate normal.

This condition imposes strong restrictions on types of hereditary mechanism models of the quantitative characters, since besides the normality of the genotypic value distributions among the individuals connected by the Pedigree, it also implies



3

Fig. 3. Double step sibs



4

Fig. 4. A relationship with one inbred relative

that all their pairwise regressions are linear and homoscedastic. This is hardly justifiable in classical genetic models of quantitative inheritance, especially with nonrandom mating (Bulmer, 1976). However, in the models of quantitative inheritance introduced by Slatkin (1970) or Cavalli-Sforza and Feldman (1976) (see also Feldman and Cavalli-Sforza, 1977, 1979; Karlin, 1979), this condition is satisfied.

Under Condition A, the genotypic covariance of relatives may be expressed as

$$\mathrm{cov}_g^{(\mathrm{REL})} = \tfrac{1}{4}(a_{13} + a_{23} + a_{14} + a_{24}), \tag{9}$$

where a_{ij} are the covariances of the distribution

$$\Pr[x_1, x_2, x_3, x_4 | \mathrm{PED}, (1*2), (3*4)]. \tag{10}$$

Before proceeding further, let us bring into consideration the following lemma introduced in the previous paper (Gimelfarb, 1981).

Lemma. *For any three events A, B, and C,*

$$\Pr[A|B, C] = \frac{\Pr[A|B]\Pr[C|A, B]}{\Pr[C|B]}. \tag{11}$$

This is easily verified by multiplying both the numerator and denominator by $\Pr[B]$.

According to the lemma, the distribution (10) can be represented as

$$\Pr[x_1, x_2, x_3, x_4 | \mathrm{PED}, (1*2), (3*4)]$$
$$= \Pr[x_1, x_2, x_3, x_4 | (1*2), (3*4)] \frac{\Pr[\mathrm{PED}|x_1, x_2, x_3, x_4, (1*2), (3*4)]}{\Pr[\mathrm{PED}|(1*2), (3*4)]}. \tag{12}$$

Given Condition H,

$$\Pr[\mathrm{PED}|x_1, x_2, x_3, x_4, (1*2), (3*4)] = \Pr[\mathrm{PED}|x_1, x_2, x_3, x_4]$$
$$= \frac{\Pr[x_1, x_2, x_3, x_4|\mathrm{PED}]\Pr[\mathrm{PED}]}{\Pr[x_1, x_2, x_3, x_4]}. \tag{13}$$

The distribution in the denominator of (13) is just the joint distribution of the genotypic values among four arbitrary individuals, two of whom are from generation n' and the other two are from generation n''. Obviously,

$$\Pr[x_1, x_2, x_3, x_4] = p_{n'}(x_1)p_{n'}(x_2)p_{n''}(x_3)p_{n''}(x_4). \tag{14}$$

Since

$$\Pr[x_1, x_2, x_3, x_4|(1*2),(3*4)] = p_{n'}(x_1, x_2|(1*2))p_{n''}(x_3, x_4|(3*4)), \tag{15}$$

combining (12), (13), (14) and (15) we obtain

$$\Pr[x_1, x_2, x_3, x_4|\text{PED},(1*2),(3*4)] = \frac{1}{Q}\frac{\Pr[x_1, x_2, x_3, x_4|\text{PED}]p_{n'}(x_1, x_2|(1*2))p_{n''}(x_3, x_4|(3*4))}{p_{n'}(x_1)p_{n'}(x_2)p_{n''}(x_3)p_{n''}(x_4)}, \tag{16}$$

where

$$Q = \frac{\Pr[\text{PED}|(1*2),(3*4)]}{\Pr[\text{PED}]} = \frac{\Pr[\text{PED},(1*2),(3*4)]}{\Pr[\text{PED}]\Pr[(1*2),(3*4)]} \tag{17}$$

is a normalizing term.

According to Conditions G and I, all the distributions in (16) are normal, i.e., are exponential functions of second degree polynomials. Since a product of exponential functions of second degree polynomials is always an exponential function of a second degree polynomial, it follows from (16) that the distribution (10) is a tetravariate normal. If **A** denotes the variance-covariance matrix of this distribution, then a_{ij} in (9) are elements of this matrix, for which the following expression emerges from (16) (see Kendall and Stuart, 1969):

$$\mathbf{A} = (\mathbf{C}_{\text{PED}}^{-1} + \mathbf{C}_{\text{MATE}}^{-1} - \mathbf{V}^{-1})^{-1}. \tag{18}$$

$\mathbf{C}_{\text{PED}}$ is the variance-covariance matrix of the distribution $\Pr[x_1, x_2, x_3, x_4|\text{PED}]$,

$$\mathbf{C}_{\text{PED}} = \begin{pmatrix} v_1 & c_{12} & c_{13} & c_{14} \\ c_{12} & v_2 & c_{23} & c_{24} \\ c_{13} & c_{23} & v_3 & c_{34} \\ c_{14} & c_{24} & c_{34} & v_4 \end{pmatrix}, \tag{19}$$

where v_i is the genotypic variance for individual i, given the Pedigree, and c_{ij} is the genotypic covariance of individuals i and j, given the Pedigree. $\mathbf{C}_{\text{MATE}}$ is the matrix

$$\mathbf{C}_{\text{MATE}} = \begin{pmatrix} v_{n'} & c_{n'}^* & 0 & 0 \\ c_{n'}^* & v_{n'} & 0 & 0 \\ 0 & 0 & v_{n''} & c_{n''}^* \\ 0 & 0 & c_{n''}^* & v_{n''} \end{pmatrix}, \tag{20}$$

where $v_{n'}, v_{n''}$ are the genotypic variances; and $c_{n'}^*, c_{n''}^*$ are the genotypic covariances of mating individuals in generations n' and n'',

$$c_{n'}^* = h_{n'}^2 \rho_{n'} v_{n'}, \tag{21a}$$

$$c_{n''}^* = h_{n''}^2 \rho_{n''} v_{n''}. \tag{21b}$$

V is the matrix of the variances $v_{n'}$ and $v_{n''}$,

$$\mathbf{V} = \begin{pmatrix} v_{n'} & 0 & 0 & 0 \\ 0 & v_{n'} & 0 & 0 \\ 0 & 0 & v_{n''} & 0 \\ 0 & 0 & 0 & v_{n''} \end{pmatrix}. \tag{22}$$

The covariances of relatives can be obtained by computing (18) and (9), given, of course, that the elements of the matrix $\mathbf{C}_{\text{PED}}$ are known. It is important to realize that elements of this matrix are defined as variances and covariances for individuals 1, 2, 3, 4 in the reproducing phase. If there is selection in the population, they may differ from the variances and covariances for the same individuals in the juvenile phase. Usually (as a result of previous computations) the matrix $\mathbf{C}_{\text{PED}}$ is known for the juvenile phase. In order to be able to use (18), it is necessary to have it transformed into the matrix for the reproducing phase. The transformation will depend, of course, on the form of selection. It can be performed analytically in the case of normalizing selection (Feldman and Cavalli-Sforza, 1977), whereas for other forms of selection an analytical solution may not be possible.

Even if the matrix $\mathbf{C}_{\text{PED}}$ for the reproducing phase is known, to obtain the elements of the matrix $\mathbf{A}$ from (18) analytically is not always an easy task, and numerical computations must be used instead. Analytical solutions can be found, however, in the cases of double first cousins and of the relationship in Fig. 4.

Double First Cousins (Fig. 2)

The elements of matrices $\mathbf{C}_{\text{PED}}$, $\mathbf{C}_{\text{MATE}}$ and $\mathbf{V}$ for this relationship are

$$\begin{aligned} v_1 &= v_2 = v_3 = v_4 = v_{n'} = v_{n''} = v_n, \\ c_{12} &= c_{14} = c_{23} = c_{34} = 0, \\ c_{13} &= c_{24} = r_{\text{sib}} v_n, \\ c^*_{n'} &= c^*_{n''} = r_n v_n, \end{aligned} \tag{23}$$

where v_n is the genotypic variance in the population in generation n; r_{sib} is the correlation coefficient of genotypic values of sibs; r_n is the correlation coefficient of genotypic values of mating individuals in generation n. The elements of matrix A computed from (18) are

$$a_{11} = a_{22} = a_{33} = a_{44} = \frac{(1 + r^2_{\text{sib}} r^2_n) v_n}{1 - r^2_{\text{sib}} r^2_n}, \tag{24a}$$

$$a_{12} = a_{34} = \frac{(1 + r^2_{\text{sib}}) r_n v_n}{1 - r^2_{\text{sib}} r^2_n}, \tag{24b}$$

$$a_{13} = a_{24} = \frac{r_{\text{sib}}(1 + r^2_n) v_n}{1 - r^2_{\text{sib}} r^2_n}, \tag{24c}$$

$$a_{14} = a_{23} = \frac{2 r_{\text{sib}} r_n v_n}{1 - r^2_{\text{sib}} r^2_n}. \tag{24d}$$

An expression for r_{sib} can be easily obtained from formula (28) in (Gimelfarb, 1981)

$$r_{\text{sib}} = \frac{1}{2}(1 + h^2_{n-1} \rho_{n-1}) \frac{v_{n-1}}{v_n}. \tag{25}$$

By combining (25), (24c), (24d) and (9), the expression for the double first cousin

genotypic covariance is obtained:

$$\mathrm{cov}_g^{(\mathrm{D.F.C.})} = \frac{1}{4} \frac{(1 + r_{n-1})(1 + r_n)^2 v_{n-1}}{1 - \frac{1}{4}(1 + r_{n-1})^2 r_n^2 \left(\frac{v_{n-1}}{v_n}\right)^2}, \tag{26}$$

where r_{n-1} is the genotypic correlation of mating individuals in generation $n-1$,

$$r_{n-1} = h_{n-1}^2 \rho_{n-1}. \tag{27}$$

At equilibrium, when $r_{n-1} = r_n = r$ and $v_{n-1} = v_n = v$, the expression for the double first cousin genotypic covariance is reduced to

$$\mathrm{cov}_g^{(\mathrm{D.F.C.})} = \frac{1}{4} \frac{(1 + r)^3}{1 - \frac{1}{4}(1 + r)^2 r^2} v. \tag{28}$$

Notice that the factor in front of v in this expression is not the genotypic correlation of double first cousins. In order to obtain this correlation, (28) ought to be divided by $v_{\mathrm{D.F.C.}}$, the genotypic variance of double first cousins. As we shall see, this variance is not the same as v, the genotypic variance in the whole population. Assuming, however, that the assortative mating is not strong, i.e., $|r| \ll 1$ and that $V = V_{\mathrm{D.F.C.}}$ (which may indeed be approximately true if assortative mating is weak), an approximate formula for the genotypic correlation emerges:

$$r_g^{(\mathrm{D.F.C.})} \approx \tfrac{1}{4}(1 + 3r). \tag{29}$$

This agrees with the formula given by Fisher (1918) for the double first cousin genotypic correlation in the case of additive characters. Notice also that the expression (28) does not agree with the formula for the double first cousin covariance presented by Bulmer (1980).

As another example of the application of expression (18), let us consider the relationship depicted in Fig. 4. This is an example of Type 2 relatives (Gimelfarb, (1981) for which Conditions A–F alone were not sufficient to obtain their covariance. This relationship is particularly interesting because one of the relatives is inbred. Whether relatives are inbred or not becomes irrelevant, however, as long as it is assumed that Conditions G, H and I hold. Therefore, the same procedure as in the case of double first cousins may be employed in order to obtain the covariance. For the elements of matrices $\mathbf{C}_{\mathrm{PED}}$, $\mathbf{C}_{\mathrm{MATE}}$ and $\mathbf{V}$ we have in this case

$$\begin{aligned} v_1 &= v_2 = v_3 = v_4 = v_{n'} = v_{n''} = v_n, \\ c_{14} &= c_{24} = c_{34} = 0, \\ c_{12} &= c_{13} = c_{23} = r_{\mathrm{h.sib}} v_n, \\ c_{n'}^* &= c_{n''}^* = r_n v_n, \end{aligned} \tag{30}$$

where v_n and r_n are the same as previously, and $r_{\mathrm{h.sib}}$ is the genotypic correlation of half sibs, for which the following expression can be derived from formula (34) in (Gimelfarb, 1981):

$$r_{\mathrm{h.sib}} = \frac{1}{4}(1 + 2h_{n-1}^2 \rho_{n-1} + h_{n-1}^2 \rho_{n-1}^2) \frac{v_{n-1}}{v_n}. \tag{31}$$

Computing the elements of matrix **A** from (18) and making use of (9) we obtain the expression for the genotypic covariance of the relatives.

$$\mathrm{cov}_g^{(\mathrm{REL})} = \frac{1}{2}\frac{r_{\mathrm{h.sib}}(1+r_n)}{1-r_{\mathrm{h.sib}}r_n}v_n. \tag{32}$$

The following questions may be asked about a relationship: (1) Is the frequency of the relationship in the population (i.e., the probability that two arbitrary chosen individuals have the given relationship) the same under assortative mating as it is under random mating? (2) Is the subpopulation of the relatives a representative sample of the whole population with respect to the quantitative character distribution, i.e., does the fact that an individual has a relative of the given type affect the probability for an individual to have a particular character?

Answers to these questions are trivial in the cases of descendant-ancestor and full sib relationships. Of course, the frequency of descendant-ancestor pairs or of full sib pairs is always the same under assortative mating as it is under random mating. Of course, the distribution of the character in a subpopulation of individuals having an ancestor or a descendant or a full sib is the same as in the whole population. In spite of the triviality of the answers in the cases of descendant-ancestor and full sib relationships, they are not always trivial for other relationships, as we shall see.

Let us have a look at the normalizing term Q in (16) for which the following expression has been obtained:

$$Q = \frac{\Pr[\mathrm{PED}, (1*2), (3*4)]}{\Pr[\mathrm{PED}]\Pr[(1*2), (3*4)]}. \tag{33}$$

Being a normalizing term, Q is computed by integrating the right side of (16) over all $\{x\}$ and making the result equal to 1. However, besides being a normalizing term, Q has an interesting meaning.

Notice, that the probabilities in the denominator of (33) remain the same whether the matings in generations n' and n'' are random or not. Indeed, the first probability, Pr[PED] has nothing to do at all with these matings, whereas $\Pr[(1*2), (3*4)]$ is the probability of matings between some unspecified individuals in the reproducing phase, which is, of course, the same for any mating system.

If matings in generations n' and n'' were random, the integration of (16) would have yielded $Q = 1$. Therefore,

$$\Pr^{(\mathrm{random})}[\mathrm{PED}, (1*2), (3*4)] = \Pr[\mathrm{PED}]\Pr[(1*2), (3*4)], \tag{34}$$

and (33) may be rewritten as

$$Q = \frac{\Pr^{(\mathrm{assort})}[\mathrm{PED}, (1*2), (3*4)]}{\Pr^{(\mathrm{random})}[\mathrm{PED}, (1*2), (3*4)]}, \tag{35}$$

where "assort" and "random" refer only to the matings exercised in generations n' and n''. If Pr[REL] denotes the frequency of the relatives in the population, i.e., the probability that two arbitrary individuals from generations $n'+1$ and $n''+1$ have the relationship specified as REL, then obviously

$$\Pr[\mathrm{REL}] = \Pr[\mathrm{PED}, (1*2), (3*4)]. \tag{36}$$

Hence,

$$Q = \frac{\Pr^{(\text{assort})}[\text{REL}]}{\Pr^{(\text{random})}[\text{REL}]}, \tag{37}$$

and Q has the following meaning. It is the ratio of the frequency of a specific relationship in the population with assortative matings in generations n' and n'' to the frequency with which this relationship would have occurred if matings in these generations were random.

Under Conditions G and I, all distributions in (16) are normal. Therefore, integration of (16) over all $\{x\}$ yields the following expression for Q (Kendall and Stuart, 1969):

$$Q = \left(\frac{\det_{\text{PED}} \det_{\text{MATE}}}{\det_V \det_A}\right)^{1/2}, \tag{38}$$

where det stands for the determinant of the corresponding matrix. It is not difficult to see from (38) that under assortative mating, Q is not necessarily equal to 1.

Let us now have a look at $\Pr[z'|\text{REL}]$, the distribution of the genotypic values in the subpopulation of the relatives in the juvenile phase. This distribution can be represented as

$$\Pr[z'|\text{REL}] = \int_{x_1}\int_{x_2} \Pr[x_1, x_2|\text{PED}, (1*2), (3*4)] H_{n'}(z'|x_1, x_2)\, dx_1\, dx_2. \tag{39}$$

By the same arguments that led to expression (16), it can be shown that

$$\Pr[z'|\text{REL}] = \frac{1}{Q}\int_{x_1}\int_{x_2} \frac{\Pr[x_1, x_2|\text{PED}]\, p_{n'}(x_1, x_2|(1*2))}{p_{n'}(x_1)\, p_{n'}(x_2)} H_{n'}(z'|x_1, x_2)\, dx_1\, dx_2. \tag{40}$$

Comparing this expression with expression (1) for the distribution of the genotypic values among the juvenile individuals in the whole population, we see that they are different. The same is true, of course, for $\Pr[z''|\text{REL}]$. Therefore, in general a subpopulation of relatives is not necessarily a representative sample of the whole population with respect to the character distribution.

Three types of relatives have been introduced in the previous paper (Gimelfarb, 1981) depending on the connections through the Pedigree. Type 1 relatives are those that are connected through only one of their parents. Relatives are of Type 2 if both parents of one of them are connected to only one of the parents of the other. Type 3 relatives are those connected through both of their parents. It has been shown that all "traditional" relationships are either of Type 1 or of a special case of Type 2, that is "Nth uncle-niece."

It is easy to see that for a Type 1 relationship,

$$\Pr[x_1, x_2, x_3, x_4|\text{PED}] = p_{n'}(x_1)\, p_{n''}(x_4) \Pr[x_2, x_3|\text{PED}]. \tag{41}$$

Substitution of (41) into (16) will yield after obvious cancellations and transformations

$$\Pr[x_1, x_2, x_3, x_4|\text{PED}, (1*2), (3*4)] = \Pr[x_2, x_3|\text{PED}]\, p_{n'}(x_1|x_2, (1*2))\, p_{n''}(x_3|x_4, (3*4)). \tag{42}$$

That $Q = 1$ in this case is easily verified by observing that integration of the right side of (42) over all $\{x\}$ gives 1. Hence, the frequency of a Type 1 relationship is the same under assortative matings in generations n' and n'' as it is under random mating in these generations. In order to answer the question whether a subpopulation of Type 1 relatives is a representative sample of the whole population with respect to the quantitative character distribution, let us compare the distributions of the genotypic values in the parental pairs of the relatives: $\Pr[x_1, x_2|\text{PED}, (1*2), (3*4)]$ and $\Pr[x_3, x_4|\text{PED}, (1*2), (3*4)]$ with the distributions of the genotypic values in the parental pairs in the whole population: $p_{n'}(x_1, x_2|(1*2))$ and $p_{n''}(x_3, x_4|(3*4))$. From Fig. 5 follows

$$\Pr[x_1, x_2|\text{PED}, (1*2), (3*4)] = \Pr[x_2|\text{PED}]\, p_{n'}(x_1|x_2, (1*2)),$$

$$\Pr[x_3, x_4|\text{PED}, (1*2), (3*4)] = \Pr[x_3|\text{PED}]\, p_{n''}(x_4|x_3, (3*4)). \tag{43}$$

If $\Pr[x_2|\text{PED}] = p_{n'}(x_2)$ and $\Pr[x_3|\text{PED}] = p_{n'}(x_3)$, then the genotypic value distributions in the parental pairs of the relatives are the same as the distributions in the parental pairs in the whole population. The conclusion, therefore, is that in the cases when the Pedigree does not affect the genotypic value distributions, a subpopulation of Type 1 relatives is a representative sample of the whole population with respect to the character distribution.

For a Type 2 relationship (Fig. 6),

$$\Pr[x_1, x_2, x_3, x_4|\text{PED}] = \Pr[x_1, x_2, x_3|\text{PED}]\, p_{n''}(x_4). \tag{44}$$

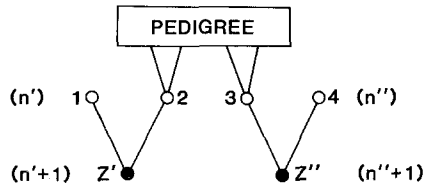
From this, according to (16) follows

$$\Pr[x_1, x_2, x_3, x_4|\text{PED}, (1*2), (3*4)]$$
$$= \frac{1}{Q} \frac{\Pr[x_1, x_2, x_3|\text{PED}]\, p_{n'}(x_1, x_2|(1*2))}{p_{n'}(x_1)\, p_{n'}(x_2)} p_{n''}(x_4|x_3, (3*4)). \tag{45}$$

Integrating this expression over all $\{x\}$ we obtain

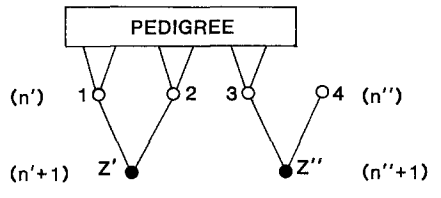
$$Q = \int_{x_1} \int_{x_2} \frac{\Pr[x_1, x_2|\text{PED}]\, p_{n'}(x_1, x_2|(1*2))}{p_{n'}(x_1)\, p_{n''}(x_2)} dx_1\, dx_2. \tag{46}$$

It is seen that Q here is not necessarily equal to 1. Therefore, the frequency of a Type 2 relationship in a population with assortative mating may differ from the frequency of this relationship in the population with random mating. Let us



5

Fig. 5. Type 1 relationship



6

Fig. 6. Type 2 relationship

assume, however, that

$$\Pr[x_1, x_2|\text{PED}] = p_{n'}(x_1)p_{n'}(x_2), \tag{47}$$

and

$$\Pr[x_3|\text{PED}] = p_{n''}(x_3). \tag{48}$$

I.e., we assume that individuals 1 and 2 have no connections through the Pedigree, and that the genotypic value distributions for individuals 1, 2, 3 are not affected by the Pedigree. Notice, that there is only one case of Type 2 relationships for which both assumptions hold. This is when individuals 1 and 2 form an Nth ancestral pair with respect to individual 3, in which case the relationship may be called "Nth uncle-niece". Given (47), it immediately follows from (46) that $Q = 1$. Moreover, since in this case

$$\Pr[x_1, x_2|\text{PED}, (1 * 2), (3 * 4)] = p_{n'}(x_1, x_2|(1 * 2)),$$

it follows also that

$$\Pr[z'|N\text{th uncle-niece}] = p_{n'+1}(z'). \tag{49}$$

And from (48) follows that

$$\Pr[z''|N\text{th uncle-niece}] = p_{n''+1}(z''). \tag{50}$$

Thus, although in general the frequency of a Type 2 relationship in a population under assortative mating may differ from the frequency of this relationship under random mating, in the case of Nth uncle-niece these frequencies are the same. Also, although in general a subpopulation of Type 2 relatives is not a representative sample of the whole population with respect to the character distribution, in the case of Nth uncle-niece it is a representative sample.

In general, in order to determine the value of Q in cases when Conditions G – I are assumed to hold, expression (38) may be used. Let us consider, for example, the case of double first cousins. For this relationship, (38) gives

$$Q^{(\text{D.F.C.})} = (1 - r_{\text{sib}}^2 r_n^2)^{-1}. \tag{51}$$

It is seen from this that under any assortative mating $Q^{(\text{D.F. Cosin})} > 1$, which means that the frequency of double first cousins in a population with assortative mating is always higher than when mating is random. In a population with heritability 0.75 and marital correlation also 0.75, for example, the value of Q will be 1.24. Hence, there will be 24% increase in the frequency of double first cousins as compared to their frequency in the same population under random mating.

For the genotypic variance among individuals 1, 2, 3, 4 in the case of double first cousins, (24a) gives

$$v^{(1,2,3,4)} = \frac{1 + r_{\text{sib}}^2 r_n^2}{1 - r_{\text{sib}}^2 r_n^2} v_n. \tag{52}$$

From this follows that under any assortative mating, the variance among parents of double first cousins is always higher than in the whole population. Therefore, it can be argued that in the subpopulation of double first cousins the variance will be

higher than in the whole population. In a population with heritability 0.75 and marital correlation 0.75, the genotypic variance among the parents of double first cousins will be 66% higher than in the whole population. Although we cannot say exactly what will be the difference of the genotypic variances in the subpopulation of double first cousins and in the whole population since it depends on properties of the function $H_n(z|x_i, x_j)$, we may expect it to be close to the same value of 66%.

Conclusion

The frequency of a relationship in a population under assortative mating may in general be different from the frequency of the same relationship in the population under random mating. Also the distribution of a quantitative character in the subpopulation of relatives is not necessarily the same as in the whole population. Therefore, methods for computing correlations between relatives that are based exclusively on linear relationships between variables (such as the path analysis) may lead to erroneous results when applied to arbitrary relationships. However, for any relationship which is a combination of descendant-ancestor, full sib, Type 1 and Nth uncle-niece relationships, its frequency in a population under assortative mating is the same as in the population under random mating, and the distribution of a quantitative character in the subpopulation of relatives is the same as in the whole population. All "traditional" relationships are included in this category, and these are the same relationships for which Conditions A – F alone are sufficient in order to obtain the covariance of the relatives (Gimelfarb, 1981). Results obtained for such relationships by general linear methods are, therefore, quite correct.

An expanded set of conditions, which besides Conditions A – F also includes Conditions G – I, enables the covariances of relatives to be computed for a larger variety of relationships.

Acknowledgements. I am indebted to James Crow and Bill Engels for important suggestions and to Professor Sewall Wright for valuable discussions. Communications with Thomas Nagylaki were especially useful. I also thank both referees for their valuable criticism.

References

Bulmer, M. G.: Regression between relatives. Genet. Res. **28**, 199 – 203 (1976)

Bulmer, M. G.: The mathematical theory of quantitative genetics. Oxford: Clarendon Press 1980

Cavalli-Sforza, L. L., Feldman, M. W.: Evolution of continuous variation: Direct approach through joint distribution of genotypes and phenotypes. Proc. Natl. Acad. Sci. USA **73**, 1689 – 1692 (1976)

Feldman, M. W., Cavalli-Sforza, L. L.: Quantitative inheritance, stabilizing selection, and cultural evolution. In: Proceedings of the international conference on quantitative genetics (Pollak, H., Kempthorne, O., eds.), pp. 761 – 777, 1977

Feldman M. W., Cavalli-Sforza, L. L.: Aspects of variance and covariance analysis with cultural inheritance. Theoret. Population Biology **15**, 276 – 307 (1979)

Fisher, R. A.: The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh **52**, 399 – 433 (1918)

Gimelfarb, A.: A general linear model for the genotypic covariance between relatives under assortative mating. J. Math. Biol. **13**, 209 – 226 (1981)

Karlin, S.: Models of multifactorial inheritance: I. Multivariate formulations and basic convergence results. Theoret. Population Biology **15**, 308–355 (1979)
Kendall, M. G., Stuart, A.: The advanced theory of statistics. New York: Hafner Publishing Company 1969
Slatkin, M.: Selection and polygenic characters. Proc. Natl. Acad. Sci. USA **66**, 87–93 (1970)

Received March 23/Revised July 20, 1981