Is offspring–midparent regression affected by assortative mating of parents?

A. GIMELFARB

Department of Biology (Whitman), The University of Chicago, Chicago, IL 60637

Summary

An analysis based on a model that is different from the traditional Fisher's model for quantitative characters under assortative mating reveals that the genotypic offspring–midparent regression can be affected by assortative mating of parents. It is demonstrated that the prediction that mating parents assortatively introduces only a negligible bias in the estimated coefficient of linear offspring-midparent regression is limited to Fisher's model and cannot be generalized.

1. Introduction

The regression of offspring on midparental value of a quantitative character has been widely used to estimate the heritability of the character in a population. Considering experimental methods of estimating the coefficient of offspring–midparent regression in a randomly mating population, Reeve (1953) suggested that 'statistically more accurate estimates can be obtained with the same sample size by mating parents assortatively' (Reeve, 1961). This suggestion is based on the fact that the variance of midparental values is increased by assortative mating of parents, and, consequently, the variance of the estimated regression coefficient is decreased. Falconer (1981) also recommended assortative mating of parents as a way to obtain a better estimate of the offspring–midparent regression for a randomly mating population.

It is clear that Reeve's method can be useful only if the assortative mating of parents does not introduce a bias in the estimated value of the regression coefficient as compared to its value in the population under random mating. Wright (1952) pointed out that correlations between non-additive components in parents arising under assortative mating may introduce a bias in the estimated regression coefficient. However, Reeve (1961) concluded that such a bias will be negligible. He drew this conclusion from the model of quantitative characters under assortative mating introduced by Fisher (1918). The main assumptions of this model pertinent to the Reeve's analysis are:

(1) The quantitative character is controlled by an infinite number of loci, each having an infinitesimal effect on the character.

(2) There are only two alleles in each locus.

(3) There is no epistasis between effects of loci.

(4) The distribution of the genotypic values of individuals in mating pairs is bivariate normal.

Clearly, these assumptions made in 1918 at the dawn of genetics are quite simplistic in view of the modern knowledge accumulated in genetics since then. It is certain that these assumptions may not always adequately reflect the complex reality of the heredity and development of quantitative characters in organisms. It is well known, for example, that a locus can have more, and sometimes much more, than just two alleles (how to define a locus controlling a quantitative character represents an additional problem). It is also known that dominance is not the only way in which genes can interact, and that epistasis is as widely present as dominance. Moreover, in some instances a distinction between dominance and epistasis is quite 'blurred' (Lerner, 1958). In view of all this, it is important to make sure that a conclusion reached for Fisher's model is not limited to this particular model, but possess a broader generality. This becomes especially important when the conclusion leads to a practical recommendation, as in the case with the suggested method of mating parents assortatively in order to obtain a statistically more accurate estimate of the offspring-midparent regression.

The purpose of this paper is to demonstrate that the conclusion that only a negligible bias is introduced in the estimated offspring–midparent regression coefficient by assortative mating of parents results from the assumptions of Fisher's model and is, therefore, limited to this model. It will be shown that, when a different model is used to describe quantitative characters under assortative mating, the model predicts that a substantial bias can be introduced in the regression coefficient by mating parents assortatively as compared to mating them randomly.

2. The model

The model that will be used in the analysis is the gametic model introduced by Gimelfarb (1982). A gamete is treated in this model as the elementary unit controlling a quantitative character and a finer genic constitution of gametes (loci, alleles) is not specified. Each gamete is characterized by a number (called *gametic contribution*), and the genotypic value, X, of an individual whose genotype is made up by gametes with contributions α and β is considered to be a function (called *developmental function*), $f\alpha, \beta$), of the gametic contributions,

$$X = f(\alpha, \beta). \tag{1}$$

The particular form of this function can be very complex for a given quantitative character, since it will depend on particularities of the genetic system controlling the character. It has been suggested, however, (Gimelfarb, 1982) that the form

$$f(\alpha, \beta) = A(\alpha + \beta) + D\alpha\beta \tag{2}$$

can serve as an approximation to an actual developmental function of a quantitative character. It is important to realize that a developmental function in the form (2) does not imply a specific gene action. It should be viewed only as a convenient mathematical approximation incorporating interactions on the gametic level. When $D = 0$, the gametic contributions are strictly additive, when $A = 0$, they are strictly multiplicative. No scale transformation exists that can convert (2) into an additive form when both parameters A and D are non-zero. Notice also that although for many quantitative characters (2) is just an approximation to the actual developmental function, there may be characters for which (2) is the exact form of their actual developmental functions.

The genotypic composition of a population is described in the gametic model by the bivariate distribution $p(\alpha, \beta)$ of the genotypes of individuals expressed in terms of gametic contributions. Due to the symmetry of gametes in genotypes, $p(\alpha, \beta) = p(\beta, \alpha)$. The univariate marginal of $p(\alpha, \beta)$, $p(\alpha) = \int p(\alpha, \beta) \, d\beta$, represents the distribution of gametic contributions in the pool of gametes of the population.

An assumption will be made that distribution $p(\alpha, \beta)$ is bivariate normal in any generation. Notice that given this assumption and the developmental function in form (2) with a non-zero D the distribution of the genotypic values for the character will not be normal. This is one of the differences between this model and the model of Fisher, where the distribution of genotypic values is assumed as normal in any generation.

Under the assumption of bivariate normality, the distribution $p(\alpha, \beta)$ is totally described by the following parameters: the mean gametic contribution, m, the variance of gametic contributions, v, and the covariance

of contributions of the gametes constituting a zygote, cov:

$$m = \int \alpha p(\alpha) \, d\alpha, \tag{3a}$$

$$v = \int \alpha^2 p(\alpha) \, d\alpha - m^2, \tag{3b}$$

$$\text{cov} = \iint \alpha\beta p(\alpha, \beta) \, d(\alpha\beta) - m^2 \tag{3b}$$

It will be assumed that parents are drawn from a population in equilibrium under random mating (including linkage equilibrium). Therefore, the covariance of gametic contributions in a parental zygote is zero. We shall also assume for simplicity that the mean contribution of gametes in parents is also zero. Thus, the parental distribution $p(\alpha, \beta)$ has the following parameters:

$$m = 0, \tag{4a}$$

$$v = v_0, \tag{4b}$$

$$\text{cov} = 0, \tag{4c}$$

where v_0 is the variance of gametic contributions in linkage equilibrium.

Given the developmental function as in (2), the mean genotypic value in parents, M_P, is computed as

$$M_P = E[f(\alpha, \beta)] = E[A(\alpha + \beta) + D\alpha\beta]$$
$$= 2Am + D(\text{cov} + m^2).$$

Because of (4a) and (4c),

$$M_P = 0. \tag{5}$$

The variance, V_P, of the genotypic values in parents is computed as follows.

$$V_P = E[f^2(\alpha, \beta)] - M_P^2$$
$$= E[A^2(\alpha + \beta)^2 + D^2\alpha^2\beta^2 + 2AD\alpha\beta(\alpha + \beta)],$$

or, taking into consideration (4a) and (4c),

$$V_P = 2A^2 v_0 + D^2 m_{22} + 4AD m_{21}, \tag{6}$$

where m_{22} and m_{21} are the higher moments of the distribution $p(\alpha, \beta)$. Due to the assumption of bivariate normality,

$$m_{21} = 0 \tag{7a}$$

$$m_{22} = v_0^2 + 2 \, \text{Cov}^2 = v_0^2 \tag{7b}$$

(Kendall & Stuart, 1973). The substitution of (7a) and (7b) into (6) yields

$$V_P = 2A^2 v_0 + D^2 v_0^2. \tag{8}$$

It is clear that, since the mean parental genotypic value, M_P, is zero, the mean midparental genotypic value, $M_{\bar{P}}$, is also zero. As for the variance of the midparental genotypic values, $V_{\bar{P}}$, it is known (Falconer, 1981) that

$$V_{\bar{P}} = \tfrac{1}{2} V_P (1 + R), \tag{9}$$

where R is the correlation between genotypic values of parents.

Considering the relationship between the gametic distributions in offspring and in parents, it is obvious that, given parental gametes are in linkage equilibrium, the gametic distribution in offspring is the same as in parents. Therefore,

$$m' = m = 0, \quad v' = v = v_0, \tag{10a,b}$$

where primes indicate offspring parameters.

We shall now make an assumption about the assortative mating of parents. Consider the distribution of the genotypes expressed in terms of gametic contributions in mating pairs, $P(\alpha, \beta; \gamma, \delta)$, representing the frequency of pairs with one mate having genotype (α, β) and the other having genotype (γ, δ). It will be assumed that this distribution is tetravariate normal. This will imply that $p(\alpha, \beta)$, which represents, of course, the bivariate marginal of P, is bivariate normal. Assuming that the quantitative character is sex-independent,

$$P(\alpha, \beta; \gamma, \delta) = P(\gamma, \delta; \alpha, \beta). \tag{11}$$

Due to the assumption of tetravariate normality, this distribution is completely determined by the four parameters:

$$m = 0, \quad v = v_0, \quad \text{cov} = 0, \quad \text{cov}^*, \tag{12}$$

where cov* is the covariance between contributions of gametes from the genotypes of two mating individuals. Notice that, although cov = 0, because of the random mating in the population from which parents have been drawn, cov* $\neq$ 0, if parents mate assortatively.

The last assumption that will be made concerns the genetic transmission of quantitative characters. We shall assume that the mean contribution of gametes produced by an individual is equal to the average of the contributions of the gametes constituting the individual's genotype. Thus, if $\mu(\alpha, \beta)$ designates the mean contribution of gametes produced by an individual whose genotype is (α, β),

$$\mu(\alpha, \beta) = \tfrac{1}{2}(\alpha + \beta). \tag{13}$$

Notice that (13) is the only assumption about the genetic transmission of a quantitative character made in this paper. This assumption always holds true if the character is controlled by very tightly linked loci (no recombination), which is equivalent to one multiallelic locus. It will also be true for any recombination pattern if, for example, the gametic contribution has a meaning of the number of positive (or negative) alleles in the gamete. In other instances (13) can be considered as an approximation.

Thus, there are three main assumptions on which the following analysis will be based:

(1) The developmental function, $f(\alpha, \beta)$, is as in (2).

(2) The distribution of gametic contributions in pairs of mating individuals, $P(\alpha, \beta; \gamma, \delta)$, is tetravariate normal.

(3) The mean contribution of a gamete produced by an individual is equal to the average contribution of the gametes constituting the individual's genotype (eq. 13).

3. Offspring–midparent genotypic regression

The coefficient of linear regression, b, of the offspring's genotypic value on the midparental genotypic value is by definition,

$$b = \frac{\text{Cov}_{O-\bar{P}}}{V_{\bar{P}}}, \tag{14}$$

where $\text{Cov}_{O-\bar{P}}$ is the covariance between the genotypic value of offspring and the midparental genotypic value. It is known (Crow & Kimura, 1970) that this covariance is the same as the covariance between the genotypic values of offspring and parent, Cov_{O-P}. Therefore, taking into account (9),

$$b = \frac{2\,\text{Cov}_{O-P}}{V_P(1+R)}. \tag{15}$$

Let (λ, η) be the genotype of an offspring, whereas (α, β) and (γ, δ) be the genotypes of parents. Then the covariance between the genotypic values of offspring and a parent can be written as

$$\begin{aligned}
\text{Cov}_{O-P} &= E\{[A(\lambda + \eta) + D\lambda\eta]\,[A(\alpha + \beta) + D\alpha\beta]\} \\
&= E\{[A(\mu(\alpha, \beta) + \mu(\gamma, \delta)) \\
&\quad + D\mu(\alpha, \beta)\,\mu(\gamma, \delta)]\,[A(\alpha + \beta) + D\alpha\beta]\},
\end{aligned}$$

where $\mu(\alpha, \beta)$ and $\mu(\gamma, \delta)$ are the mean contributions of gametes produced by parents with genotypes (α, β) and (γ, δ), respectively. Taking into consideration (13),

$$\begin{aligned}
\text{Cov}_{O-P} &= \tfrac{1}{2}E\{A^2(\alpha + \beta)(\alpha + \beta + \gamma + \delta) \\
&\quad + \tfrac{1}{2}D^2\alpha\beta(\alpha + \beta)(\gamma + \delta) \\
&\quad + AD[\alpha\beta(\alpha + \beta + \gamma + \delta) + \tfrac{1}{2}(\alpha + \beta)^2\,(\gamma + \delta)]\},
\end{aligned}$$

which yields after some transformations

$$\begin{aligned}
\text{Cov}_{O-P} &= A^2(v_0 + \text{cov} + 2\,\text{cov}^*) + D^2 m_{2110} \\
&\quad + AD(m_{2100} + m_{2010} + m_{1110}), \tag{16}
\end{aligned}$$

where

$$m_{ijkl} = \iiint \alpha^i\beta^j\gamma^k\delta^l P(\alpha, \beta; \gamma, \delta)\,d(\alpha\beta\gamma\delta)$$

is the respective moment of distribution P. Due to the assumption of tetravariate normality,

$$m_{2100} = m_{2010} = m_{1110} = 0, \tag{17a}$$

$$m_{2110} = v_0\,\text{cov}^* + 2\,\text{cov}\,\text{cov}^* \tag{17b}$$

(Kendall & Stuart, 1973). Substituting (17a) and (17b) into (16) and taking into consideration (4c),

$$\text{Cov}_{O-P} = A^2 v_0 + 2A^2\,\text{cov}^* + D^2 v_0\,\text{cov}^*.$$

Introducing ρ^* as the correlation between the contributions of gametes from the genotypes of two mating individuals, so that $\text{cov}^* = \rho^* v_0$,

$$\text{Cov}_{O-P} = A^2 v_0 + (2A^2 v_0 + D^2 v_0^2)\,\rho^*. \tag{18}$$

Notice that the expression in parentheses represents the genotypic parental variance, V_P, (eq. 8). Therefore,

$$\text{Cov}_{O-P} = A^2 v_0 + V_P\,\rho^*, \tag{19}$$

and, by substituting this into (14), we obtain

$$b = \frac{2A^2 v_0 + V_P \rho^*}{V_P(1 + R)}.$$ (20)

Let b_0 be the offspring–midparent regression under random mating of parents. It follows from (20) that

$$b_0 = \frac{2A^2 v_0}{V_P}.$$ (21)

Then expression (20) for the offspring–midparent regression under assortative mating of parents, b, can be presented as

$$b = b_0 + \frac{2\rho^* - b_0 R}{1 + R}.$$ (22)

In order to be able to use this formula for evaluating numerically the difference between b and b_0, it is necessary to determine values of ρ^* corresponding to a particular R. To do that, let us consider the genotypic covariance, $\mathrm{Cov}_{P-P'}$, between mating parents.

$$\mathrm{Cov}_{P-P'} = E\{[A(\alpha+\beta)+D\alpha\beta][A(\gamma+\delta)+D\gamma\delta]\}$$
$$= 2A^2 \mathrm{cov}^* + D^2 m_{1111} + 4AD\, m_{1110},$$

where m_{1111} and m_{1110} are the higher moments of $P(\alpha, \beta; \gamma, \delta)$. Because of the assumption of tetravariate normality,

$$m_{1110} = 0,$$ (23a)

$$m_{1111} = \mathrm{cov}^2 + 2(\mathrm{cov}^*)^2 = 2(\mathrm{cov}^*)^2 \text{ (since cov} = 0)$$ (23b)

Therefore,

$$\mathrm{Cov}_{P-P'} = 4A^2 \mathrm{cov}^* + 2D^2(\mathrm{cov}^*)^2$$
$$= 4A^2 v_0 \rho^* + 2D^2 v_0^2 (\rho^*)^2.$$ (24)

On the other hand, by definition,

$$\mathrm{Cov}_{P-P'} = R V_P,$$

and, hence,

$$R V_P = 4A^2 v_0 \rho^* + 2D^2 v_0^2 (\rho^*)^2,$$

or

$$\frac{2D^2 v_0^2}{V_P}(\rho^*)^2 + \frac{4A^2 v_0}{V_P}\rho^* - R = 0.$$ (25)

It follows from (8) that

$$D^2 v_0^2 = V_P - 2A^2 v_0.$$

Substituting this into (25) and taking into account (21) yields the following quadratic equation for ρ^*:

$$2(1 - b_0)(\rho^*)^2 + 2b_0 \rho^* - R = 0.$$ (26)

The two equations (22) and (26) make it possible to evaluate the offspring–midparent genotypic regression under assortative mating of parents with a given genotypic correlation of mates, R, and a given offspring–midparent regression under random mating of parents, b_0.

It should be noted that the two assumptions:

tetravariate normality of $P(\alpha, \beta; \gamma, \delta)$ and the developmental function in form (2) impose constraints on negative values of R feasible in a population. These constraints and the fact that a negative root of (26) represents a spurious correlation between the contributions of gametes under positive assortative mating have been discussed elsewhere (Gimelfarb, 1985a). In order to avoid these complications we shall limit our consideration to only positive assortative mating, i.e. positive values of R, and to positive roots of (26).

Table 1. *Effect of assortative mating in parents on the offspring–midparent regression*

R	b_0:0·90	0·75	0·50	0·25	0·00
0·1	0·92	0·80	0·62	0·48	0·41
0·2	0·93	0·84	0·70	0·60	0·53
0·3	0·94	0·87	0·76	0·67	0·60
0·4	0·95	0·88	0·79	0·72	0·64
0·5	0·96	0·90	0·82	0·74	0·67
0·6	0·96	0·91	0·84	0·76	0·68
0·7	0·97	0·92	0·85	0·78	0·70
0·8	0·97	0·93	0·86	0·79	0·70
0·9	0·97	0·93	0·87	0·79	0·71
1·0	0·97	0·94	9·87	0·79	0·71

Table 1 shows the offspring–midparent regression coefficients computed from equations (22) and (26) for different values of genotypic correlation of mates, R, and different values of the regression coefficient under random mating of parents, b_0 (in the first row). Thus, column 2, for example, shows values of b starting from $b_0 = 0·9$ to value of $b = 0·97$ corresponding to assortative mating of parents with $R = 1$.

4. Discussion

Table 1 convincingly demonstrates that the genotypic offspring–midparent regression can be affected by assortative mating in parents. Not surprisingly, the differences between b_0 and b are not large when the value of b_0 is close to 1, i.e. gametes have mostly additive effect on the character. The differences can be substantial, however, even for a modest degree of assortative mating, for increasing deviations from additivity. If, for example, the offspring-midparent regression in random mating equilibrium is $b_0 = 0·5$, assortative mating of parents with genotypic correlation $R = 0·5$ will increase the regression to the value $b = 0·82$. It should be noted that, unlike assortative mating occurring in natural populations which is usually weak, assortative mating in an experiment, where the experimenter himself matches parents, can be quite strong. It is interesting to notice that in an extreme case of developmental function in form (2) with $A = 0$, i.e. when gametic contributions are strictly multiplicative, the offspring–midparental regression in random mating equilibrium is zero. However,

assortative mating of parents will yield in this case a non-zero regression that can take quite large values.

It should be pointed out that values of regression coefficients in Table 1 are computed for a specific set of assumptions, e.g. that the developmental function is in form (2), that the distribution $P(\alpha, \beta; \gamma, \delta)$ is tetravariate normal. Therefore, the numerical values in this table cannot be viewed as representing a general situation. Deviations from the assumptions can affect the actual numerical values of the regression coefficients. In some instances the value of the regression coefficient under assortative mating can be smaller than in Table 1, in other instances it may be even greater than in the table. The real significance of Table 1 is that it demonstrates that a substantial bias *can* be introduced in the estimated offspring–midparent regression coefficient by mating parents assortatively.

Another result based on Fisher's model was obtained by Vetta (1976). It predicts that dominance variance in a population under random mating is the same as in the population under assortative mating. That this result is also limited to the Fisher's model and does not hold for the gametic model, has been demonstrated elsewhere (Gimelfarb, 1985b).

When speculating why the discrepancy exists between the result obtained by Reeve (1961) based on Fisher's model and the result obtained in this paper, one should keep in mind that these results concern the coefficient of *linear* regression, even though the actual offspring–midparent regression can be *non-linear*. The main feature of Fisher's model, which seems to be crucial for Reeve's result, is that the offspring–midparent regression in this model is always linear (or approximately linear). This linearity is implied, at least partially, by the assumption that a character is controlled by *infinitely* many *non-interacting* loci. It is known, however, that non-additive genic effects, such as dominance (Bulmer, 1980; Gimelfarb, 1985a) and epistasis, can introduce a strong non-linearity in offspring–midparent regressions. It is clear that, even though assortative mating of parents may have no effect on the coefficient of linear regression if the regression is in fact linear, it may still have a substantial effect on the coefficient of linear regression if the actual offspring–midparent regression is non-

linear. It would seem, therefore, that the discrepancy between the results predicted by the two models can be attributed to the neglected non-linearity of offspring–midparent regression in Fisher's model used by Reeve.

Thus, the main conclusion of this paper is that assortative mating of parents, while reducing the sampling variance of the estimated value of offspring–midparent regression, can at the same time introduce a substantial bias in the estimated value.

References

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Characters.* Oxford University (Claredon) Press.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics.* New York: Harper & Row.

Falconer, D. S. (1981). *Introduction to Quantitative Genetics.* New York: The Ronald Press.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.

Gimelfarb, A. (1982). Quantitative character dynamics: gametic model. *Theoretical Population Biology* **22**, 324–346.

Gimelfarb, A. (1985a). Offspring-parent regression: How linear is it? (Submitted to *Biometrics*.)

Gimelfarb, A. (1985b). Non-additive gametic variance under assortive mating. (Submitted.)

Kendall, M. G. & Stuart, A. (1973). *The Advanced Theory of Statistics,* 3rd edn. New York: Hafner.

Lerner, I. M. (1958). *The Genetic Basis of Selection.* New York: John Wiley.

Reeve, E. C. R. (1953). Studies in quantitative inheritance. III: Heritability and genetic correlation in progeny tests using different mating systems. *Journal of Genetics* **51**, 520–542.

Reeve, E. C. R. (1961). A note on non-random mating in progeny tests. *Genetical Research* **2**, 195–203.

Vetta, A. (1976). Dominance variance in Fisher's model of assortative mating. *Annals of Human Genetics* **39**, 447–453.

Wright, S. (1952). The genetics of quantitative variability. In *Quantitative Inheritance,* ed. E. C. R. Reeve and C. H. Waddington, pp. 5–41. Her Majesty's Stationery Office, London.