The additive polygenic model with assortative mating and shared parent-offspring environment

Camille Noûs[1], Hervé Perdry[2]

[1]Laboratoire Cogitamus
camille.nous@cogitamus.fr

[2]CESP Inserm, U1018, UFR Médecine, Université Paris-Saclay, Villejuif, France
herve.perdry@universite-paris-saclay.fr

Abstract

The consequences of assortative model in the additive polygenic model have been extensively explored by several authors. In this note we extend their results by introducing a correlation between parental and offspring environments.

1 Introduction

In this note we assume that the reader is familiar with the additive polygenice model in quantitative genetics, as described in textbooks, e.g [3], [1]. The consequences of assortative mating were considered by Fisher in the very paper in which he proposed the polygenic model [4]. Many other authors have considered the same problem, using different computational approaches, in particular Wright [6], Crow and Felsenstein [2], Nagylaki [5] (see this last paper for a more extensive bibliography).

While assortative mating creates a correlation between mates' environments, the classical approach consider that the offspring environment is independent from the parental environments. Here we introduce in the model a (positive) correlation between the environmental component of the offspring, and each of its parents environments.

Our results are summarized in the following theorem.

Theorem 1 *Consider a phenotype determined by $P = A + E$, where A is the sum of independent genetic effects, and E is the environmental effect. Consider a couple with two parents indexed by 1 and 2 and an offspring indexed by 3. The values of P, A and E for these individuals are denoted P_i, A_i, E_i ($i = 1,2,3$). Denote $\mathrm{var}(P) = \sigma^2$, $\mathrm{var}(A) = a^2$ and $\mathrm{var}(E) = e^2$.*

If one assumes that the population is panmictic and that all the environment terms are independent from each other and from the A_i ; then $a^2 = 2g_0^2$ where g_0^2 is the gametic variance, and $\sigma^2 = a^2 + e^2$; the correlations between parental and offspring genetic effects are $r_{A_1A_3} = r_{A_1A_2} = \frac{1}{2}$.

We will assume on the contrary that mate choice is partially based on the value of P, so that there is a correlation $r = r_{P_1P_2}$ between the two members of a couple. We also assume that there is a correlation $\nu = r_{E_1E_3} = r_{E_2E_3}$ between the parental and offspring environment, which can be attributed for example to cultural transmission.

When r is non-zero, then there is a non-zero gametic correlation r_g between parental gametes; when ν is non-zero, there is a non-zero correlation $\rho = r_{A_i E_i}$ between the genetic and environmental components. Assuming that the population is at equilibrium, that is, that r_g and ρ are constant accross generations, one can compute r_g and ρ from r, ν, g_0^2 and e^2.

The presence of parental gametic correlation r_g induces, at equilibrium, a gametic disequilibrium between the causal loci, leading to a gametic variance

$$g^2 \simeq \frac{g_0^2}{1-r_g}.$$

The above approximation is valid unless r_g is close to 1, which in turn occurs only when the trait is almost entirely genetically determined, and when the parental correlation r is close to 1.

The additive genetic variance is then $a^2 = (1+r_g)2g^2 = \frac{1+r_g}{1-r_g}2g_0^2$. The correlations between A_i and E_j terms are summarized in the following two tables, which allow to compute the parent-offspring phenotypic correlation.

	A_1	E_1	A_2	E_2
A_1	1	ρ	$\frac{r}{\sigma^2}(a+\rho e)^2$	$\frac{r}{\sigma^2}(\rho a+e)(\rho e+a)$
E_1	ρ	1	$\frac{r}{\sigma^2}(\rho a+e)(\rho e+a)$	$\frac{r}{\sigma^2}(\rho a+e)^2$
A_2	$\frac{r}{\sigma^2}(a+\rho e)^2$	$\frac{r}{\sigma^2}(\rho a+e)(\rho e+a)$	1	ρ
E_2	$\frac{r}{\sigma^2}(\rho a+e)(\rho e+a)$	$\frac{r}{\sigma^2}(\rho a+e)^2$	ρ	1

Table 1: Correlations between A and E terms for the two parents

	A_3	E_3
A_i	$\frac{1}{2}(1+\frac{r}{\sigma^2}(a+\rho e)^2)$	ρ
E_i	$\frac{1}{2}(\rho+\frac{r}{\sigma^2}(a+\rho e)(\rho a+e))$	ν

Table 2: Correlations between A and E terms for parent / offspring pairs. Here $i = 1$ or 2.

The proof of the theorem is the object of the following sections. In section 2 we will derive the consequences of the existence of a gametic correlation r_g. In section 3, we consider the components of var(P) in an individual in a population in which there is a gametic correlation, as well as a gene-environment correlation. In section 4, we consider a couple with assortative mating and prove the results that are summarized in table 1. In the section 5, we will show how the gametic correlation r_g can be derived from a, ρ and e, we derive the results summarized in table 2, and we show how, assuming population equilibrium, ρ and ν are linked. Section 6 gives some technical details about how to compute ρ and a, given g_0, e, r and ν.

2 Gametes

Assume there are N causal loci, with N a very large number. The effect of a (haploid) gamete is $G = \sum_{i=1}^{N} g_i$, where g_i is the infinitesimal effect of locus i. The genetic effect for an individual can be decomposed in $A = G^p + G^m$, where G^p (respectively G^m) is the effect of the paternal (respectively maternal) gamete.

2.1 Inter loci correlation

Let $\mathrm{var}(g_i) = \frac{1}{N} g_0^2$ and $\mathrm{cov}(g_i, g_j) = \frac{1}{N} g_0^2 \delta(i,j)$ (for $i \neq j$). If one has $\delta(i,j) = 0$ for all $i \neq j$, then $\mathrm{var}(G) = g_0^2$. In general,

$$\mathrm{var}(G) = \sum_{i=1}^{N} \mathrm{var}(g_i) + \sum_{i \neq j} \mathrm{cov}(g_i, g_j)$$
$$g^2 = g_0^2 \left(1 + (N-1)\overline{\delta}\right)$$

where $\overline{\delta} = \frac{1}{N(N-1)} \sum_{i \neq j} \delta(i,j)$. Note that the $\delta(i,j)$ are infinitesimal (they have order of magnitude $\frac{1}{N}$).

2.2 Méiosis

Consider an individual with genetic component $A = G^p + G^m$. One has $G^p = \sum_{i=1}^{N} g_i^p$, and $G^m = \sum_{i=1}^{N} g_i^m$. A simple way to model the meiosis is to let $I_p \cup I_m = [1, \ldots, N]$ with $|I_p| = |I_m| = \frac{N}{2}$. The gamete transmitted has total effect

$$G = \sum_{i \in I_p} g_i^p + \sum_{i \in I_m} g_i^m.$$

By letting

$$H = \sum_{i \in I_p} g_i^m + \sum_{i \in I_m} g_i^p,$$

one gets $A = G + H$; H can be seen as the "untransmitted gamete".

2.3 Evolution of the inter-loci correlations $\delta_t(i,j)$

Here we consider the evolution between generation t and $t+1$, assuming that there is a correlation $r_g = \mathrm{cor}(G^p, G^m)$ between the parental gametes.

At generation t, individuals are formed by gametes with variance $g_t^2 = g_0^2 \left(1 + (N-1)\overline{\delta_t}\right)$.

The two gametes emitted by the two members of a couple in generation $t-1$ have covariance $r_g g_{t-1}^2$. Thus, an individual at generation t is formed by two gametes G^p and G^m with covariance $\mathrm{cov}(G^p, G^m) = r_g g_{t-1}^2$. Now $\mathrm{cov}(G^p, G^m) = \sum_{i,j} \mathrm{cov}(g_i^p, g_j^m)$, and all the terms in this sum play symmetric roles, hence

$$\mathrm{cov}(g_i^p, g_j^m) = \frac{1}{N^2} \mathrm{cov}(G^p, G^m) = \frac{1}{N^2} r_g g_0^2 \left(1 + (N-1)\overline{\delta_t}\right).$$

Next, let us consider a gamete emitted by this individual. Let I_p and I_m be as in subsection 2.2, and let $G = \sum_{i=1}^{N} g_i$ where $g_i = g_i^p$ if $i \in I_p$ and $g_i = g_i^m$ if $i \in I_m$.

If $i, j \in I_p$, then $g_i = g_i^p$ et $g_j = g_j^p$ and $\mathrm{cov}(g_i, g_j) = \mathrm{cov}(g_i^p, g_j^p)$, thus $\delta_{t+1}(i,j) = \delta_t(i,j)$. The same holds if $i, j \in I_m$. The remaining cases are $i \in I_p, j \in I_m$ and $i \in I_m, j \in I_p$. In both cases, $\mathrm{cov}(g_i, g_j) = \mathrm{cov}(g_i^p, g_j^m) = \mathrm{cov}(g_i^m, g_j^p)$, hence

$$\delta_{t+1}(i,j) = \frac{1}{N} r_g \left(1 + (N-1)\overline{\delta_t}\right).$$

Hence the distribution of $\delta_{t+1}(i,j)$ is a mixture between $\delta_t(i,j)$ and this term. We have

$$\overline{\delta}_{t+1} = \frac{1}{2}\delta_t + \frac{1}{2}\frac{1}{N} r_g \left(1 + (N-1)\overline{\delta_t}\right)$$

hence at equilibrium, $\overline{\delta}_t = \overline{\delta}_{t+1} = \overline{\delta}$ and

$$\overline{\delta} = \frac{1}{2}\overline{\delta} + \frac{1}{2}\frac{1}{N} r_g \left(1 + (N-1)\overline{\delta}\right)$$

from which one gets

$$\overline{\delta} = \frac{r_g}{N(1-r_g)+r_g}.$$

The mean inter-loci correlation is infinitesimal (of order $1/N$), unless r_g is very close to one. One gets

$$\begin{aligned} g^2 &= g_0^2 \left(1 + (N-1)\overline{\delta}\right) \\ &= g_0^2 \left(\frac{1}{1-\left(1-\frac{1}{N}\right) r_g}\right) \end{aligned}$$

And if $\frac{r_g}{N} \ll 1 - r_g$, which is the case again unless r_g is very close to one,

$$g^2 \simeq \frac{g_0^2}{1-r_g} \tag{1}$$

The correlation between parental gametes increases the gametic variance, by inducing an infinitesimal correlation between all the causal loci. Moreover, one has

$$\begin{aligned} \mathrm{var}(A) &= \mathrm{var}(G^p + G_m) \\ a^2 &= 2g^2 + 2\,\mathrm{cov}(G_p, G_m) \\ &= 2g^2 + 2r_g g^2 \\ &= (1+r_g) \times 2g^2 \end{aligned}$$

and finally

$$a^2 = \frac{1+r_g}{1-r_g} \times 2g_0^2 \tag{2}$$

The additive genetic value of an individual is further increased by a factor $(1+r_g)$.

2.4 Correlation between transmitted and "untransmitted" gametes

We consider a population at equilibrium. As in subsection 2.2, an individual with additive genetic value $A = G^p + G^m$ transmits a gamete $G = \sum_{i\in I_p} g_i^p + \sum_{i\in I_m} g_i^m$, and $H = \sum_{i\in I_p} g_i^m + \sum_{i\in I_m} g_i^p$ is the untransmitted gamete. Then $A = G + H$ and G et H play symmetric roles.

The gametic variance is $\mathrm{var}(G) = \mathrm{var}(H) = g^2$, and $\mathrm{cov}(G,H) = \frac{1}{2}a^2 - g^2 = g^2(1+r_g) - g^2 = g^2 r_g$, hence the correlation between G and H is $r_{GH} = r_g$.

Moreover, one has $\mathrm{cov}(A,G) = \mathrm{cov}(G+H,G) = (1+r_g)g^2$; the distribution of G conditional to A has expected value $\frac{1}{2}A$ and variance $\frac{1}{2}g^2(1-r_g)$.

3 A single individual in the population

We consider a population at equilibrium, in which there is a correlation $\rho = \mathrm{cor}(A,E) = r_{AE}$ between A and E. We will derive expressions for r_{AP} and r_{EP} that will be usefull later.

We have

$$\mathrm{var}(P) = \sigma^2 = a^2 + 2\rho ae + e^2.$$

Hence

$$\begin{aligned}\mathrm{cov}(A,P) &= \mathrm{cov}(A, A+E)\\ &= \mathrm{var}(A) + \mathrm{cov}(A,E)\\ &= a^2 + \rho ae\end{aligned}$$

and

$$r_{AP} = \frac{1}{\sigma}(a + \rho e) \tag{3}$$

In the same way,

$$\begin{aligned}\mathrm{cov}(E,P) &= \mathrm{cov}(E, A+E)\\ &= \mathrm{cov}(A,E) + \mathrm{var}(E)\\ &= \rho ae + e^2\end{aligned}$$

and

$$r_{EP} = \frac{1}{\sigma}(\rho a + e) \tag{4}$$

4 A couple with assortative mating

In this section and the following, we use the same computational approach than Nagylaki [5]. Consider the two members of a couple, indexed by $i = 1,2$, with phenotypic correlation $r_{P_1P_2} = r$. We still have $r_{A_iE_i} = \rho$ for $i = 1,2$.

It is assumed that the mate choice is based only on the phenotype. This means that, conditionnal to P_2, P_1 and A_2 are independant, that is $r_{P_1A_2\cdot P_2} = 0$ (where $r_{P_1A_2\cdot P_2}$ denotes the correlation between P_1 and A_2 conditional to P_2) ; and also $r_{A_1A_2\cdot P_2} = 0$, $r_{P_1E_2\cdot P_2} = 0$, $r_{E_1E_2\cdot P_2} = 0$.

Then, using lemma 1 (appendix A) and $r_{P_1A_2\cdot P_2} = 0$, we have

$$\mathrm{cor}(P_1, A_2) = r_{P_1A_2} = r_{P_1P_2}r_{A_2P_2}.$$

We have (equation 3) $r_{A_2P_2} = \frac{1}{\sigma}(a + \rho e)$, hence

$$r_{P_1A_2} = \frac{r}{\sigma}(a + \rho e).$$

The two members of the couple play symmetric roles so one has $r_{A_1P_2} = \frac{r}{\sigma}(a + \rho e)$.

Applying again lemma 1 to $r_{A_1A_2\cdot P_2} = 0$, we have

$$r_{A_1A_2} = r_{A_1P_2}r_{A_2P_2} = \frac{r}{\sigma^2}(a + \rho e)^2.$$

From $r_{P_1E_2\cdot P_2} = 0$ we get

$$r_{P_1E_2} = r_{P_1P_2}r_{E_2P_2} = \frac{r}{\sigma}(\rho a + e).$$

The symmetry argument gives $r_{E_1P_2} = \frac{r}{\sigma}(\rho a + e)$.

And from $r_{E_1E_2\cdot P_2} = 0$, we get

$$r_{E_1E_2} = r_{E_1P_2}r_{E_2P_2} = \frac{r}{\sigma^2}(\rho a + e)^2.$$

Then we have

$$\begin{aligned}
\operatorname{cov}(P_1, E_2) &= \operatorname{cov}(A_1 + E_1, E_2) \\
r_{P_1E_2}e\sigma &= \operatorname{cov}(A_1, E_2) + \operatorname{cov}(E_1, E_2) \\
r(\rho a + e)e &= r_{A_1E_2}ae + r_{E_1E_2}e^2 \\
r(\rho a + e) &= r_{A_1E_2}a + \frac{r}{\sigma^2}(\rho a + e)^2 e
\end{aligned}$$

hence

$$\begin{aligned}
r_{A_1E_2}a &= \frac{1}{\sigma^2}r(\rho a + e)(\sigma^2 - (\rho a + e)e) \\
&= \frac{1}{\sigma^2}r(\rho a + e)(\rho ae + a^2)
\end{aligned}$$

and finally

$$r_{A_1E_2} = \frac{r}{\sigma^2}(\rho a + e)(\rho e + a)$$

which is also the value of $r_{E_1A_2}$.

We have proved all the correlations of table 1.

5 Offsprings in assortative mating

5.1 Gametic correlation

Assume the individuals $i = 1, 2$ emit gametes G_i, with $A_i = G_i + H_i$, as in subsection 2.2. We will link $\operatorname{cor}(A_1, A_2) = r_{A_1A_2}$ with the gametic correlation $r_g = \operatorname{cor}(G_1, G_2)$.

We write

$$\operatorname{cov}(A_1, A_2) = \operatorname{cov}(G_1, G_2) + \operatorname{cov}(G_1, H_2) + \operatorname{cov}(H_1, G_2) + \operatorname{cov}(H_1, H_2).$$

All four terms on the right hand side are equal, and $\operatorname{cov}(A_1, A_2) = r_{A_1A_2}a^2$, so

$$\operatorname{cov}(G_1, G_2) = \frac{1}{4}r_{A_1A_2}a^2.$$

One has $a^2 = 2(1 + r_g)g^2$ and $\operatorname{cov}(G_1, G_2) = r_g g^2$, hence $r_g = \frac{1}{2}r_{A_1A_2}(1 + r_g)$ and finally

$$r_g = \frac{r_{A_1A_2}}{2 - r_{A_1A_2}}.$$

Equivalently,

$$r_{A_1A_2} = 2\frac{r_g}{1+r_g}.$$

Equation 2 $a^2 = \frac{1+r_g}{1-r_g}2g_0^2$ can be written

$$a^2 = \frac{2g_0^2}{1-r_{A_1A_2}}.$$

This result is identical to what is derived, for example, in [2]. The approximations made to express g^2 and, consequently, a^2, were valid unless r_g was very close to one. Here we see that this happens only in the marginal case were $r_{A_1A_2}$ is very close to one, which in turns implies r close to one and $e^2 \ll \sigma^2$: the assortative mating is almost complete, and the environment has almost no part in the trait.

5.2 Genetic component

We have $a^2 = \mathrm{var}(A_1) = \mathrm{cov}(A_1, G_1) + \mathrm{cov}(A_1, H_1)$; the two terms on the right hand side are equal, so $\mathrm{cov}(A_1, G_1) = \frac{1}{2}a^2$. In the same way, from $\mathrm{cov}(A_1, A_2) = \mathrm{cov}(A_1, G_2) + \mathrm{cov}(A_1, H_2)$ we get $\mathrm{cov}(A_1, G_2) = \frac{1}{2}\mathrm{cov}(A_1, A_2) = \frac{1}{2}r_{A_1A_2}a^2$.

The additive genetic component of the offspring is $A_3 = G_1 + G_2$ (one can also write $A_3 = G_3^p + G_3^m$ with $G_3^p = G_1$, $G_3^m = G_2$). Hence

$$\mathrm{cov}(A_1, A_3) = \mathrm{cov}(A_1, G_1) + \mathrm{cov}(A_1, G_2) = \frac{1}{2}(1 + r_{A_1A_2})a^2$$

and

$$r_{A_1A_3} = \frac{1}{2}(1 + r_{A_1A_2}).$$

Of course, the two parents play symmetric role and $r_{A_2A_3} = r_{A_1A_3}$.

5.3 Environmental component

Let us formalize our assumptions on environmental correlations. We denote $\nu = r_{E_1,E_3} = r_{E_2,E_3}$. All variables have expected value 0, and $\mathrm{var}(E_i) = e^2$ for $i = 1,2,3$. The variance-covariance matrix of (E_1, E_2, E_3) is

$$e^2\begin{pmatrix} 1 & r_{E_1E_2} & \nu \\ r_{E_1E_2} & 1 & \nu \\ \nu & \nu & 1 \end{pmatrix}$$

More over, conditionnaly to (E_1, E_2), E_3 is independent from A_1 and A_2.

Note that ν can't be too high; for example if $r_{E_1E_2} = 0$, the environment of the offspring can't be identical to both E_1 and E_2 ; the maximum value in that case is $\nu = \frac{1}{\sqrt{2}}$. In general, the above matrix is positive definite if

$$\nu \leq \sqrt{\frac{1 + r_{E_1E_2}}{2}}. \tag{5}$$

The value $\nu = \sqrt{\frac{1+r_{E_1E_2}}{2}}$ corresponds to the extreme case in which the offspring's environment is entirely determined by its parents' environments.

Let us compute $\mathrm{cov}(E_1, A_3) = \mathrm{cov}(E_1, G_1) + \mathrm{cov}(E_1, G_2)$. We have $\mathrm{cov}(E_1, A_1) = \mathrm{cov}(E_1, G_1) + \mathrm{cov}(E_1, H_1)$; the two terms in the right hand side are equal, so $\mathrm{cov}(E_1, G_1) = \frac{1}{2}\mathrm{cov}(E_1, A_1)$. In the same way, $\mathrm{cov}(E_1, G_2) = \frac{1}{2}\mathrm{cov}(E_1, A_2)$. Hence

$$\begin{aligned}\mathrm{cov}(E_1, A_3) &= \frac{1}{2}\mathrm{cov}(E_1, A_1) + \frac{1}{2}\mathrm{cov}(E_1, A_2)\\ &= \frac{1}{2}(\rho + r_{E_1A_2})ea\\ &= \frac{1}{2}(\rho + \frac{r}{\sigma^2}(\rho a + e)(a + \rho e))ea\end{aligned}$$

One has $\mathrm{cov}(A_3, E_3) = \mathrm{cov}(G_1, E_3) + \mathrm{cov}(G_2, E_3) = \frac{1}{2}\mathrm{cov}(A_1, E_3) + \frac{1}{2}\mathrm{cov}(A_2, E_3)$. As the two parents play symmetric roles, $\mathrm{cov}(A_1, E_3) = \mathrm{cov}(A_2, E_3)$ and we get $\mathrm{cov}(A_3, E_3) = \mathrm{cov}(A_1, E_3)$, hence $r_{A_1E_3} = r_{A_3E_3} = \rho$, the population beeing at equilibrium.

We have proved all the results summarized in table 2.

5.4 Linking ρ and ν

We are going to compute $r_{A_1E_3}$ in an other way, linking to ν. From the assumption of population equilibrium, we derived $r_{A_1E_3} = \rho$; the computation here is the key result allowing to compute the value ρ from ν and the other parameters of the model.

Conditionnaly to (E_1, E_2), A_3 and E_3 are independent. We are going to apply lemma 2 (appendix A), with $X = (A_1, E_3)$ and $Y = \frac{1}{e}(E_1, E_2)$. One has

$$\mathrm{var}(Y) = \begin{pmatrix} 1 & r_{E_1E_2} \\ r_{E_1E_2} & 1 \end{pmatrix}, \ \mathrm{cov}(X, Y) = \begin{pmatrix} \rho a & r_{A_1E_2}a \\ \nu e & \nu e \end{pmatrix}.$$

From lemma 2, we get

$$\mathrm{cov}(A_1, E_3) = \frac{(\rho + r_{A_1E_2})a}{1 + r_{E_1E_2}}\nu e,$$

hence

$$r_{A_1E_3} = \nu\frac{\rho + \frac{r}{\sigma^2}(\rho a + e)(\rho e + a)}{1 + \frac{r}{\sigma^2}(\rho a + e)^2}.$$

Hence, we have

$$\rho = \nu\frac{\rho + \frac{r}{\sigma^2}(\rho a + e)(\rho e + a)}{1 + \frac{r}{\sigma^2}(\rho a + e)^2},$$

which can be rewritten as

$$\nu = \rho\frac{\sigma^2 + r(\rho a + e)^2}{\rho\sigma^2 + r(\rho a + e)(\rho e + a)}.$$

In the next section we show how this relation can be used to compute ρ, knowing ν, g_0 and e.

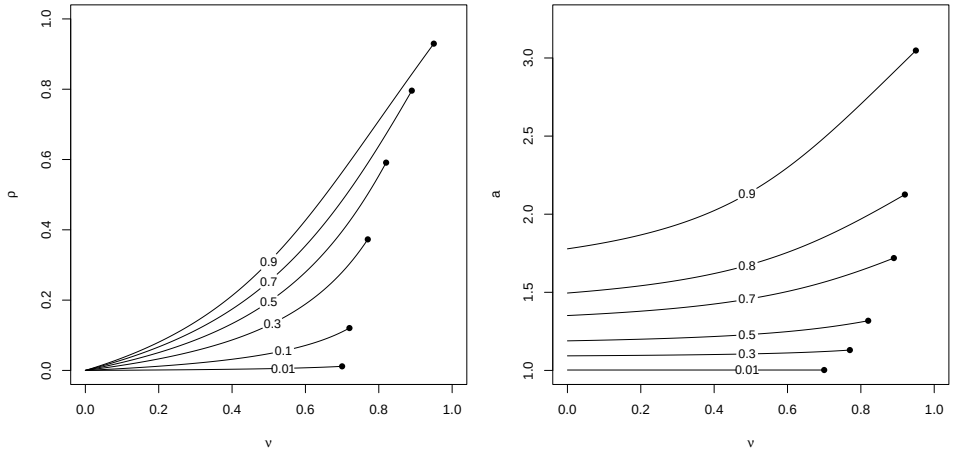

Figure 1: Relation between v and a, ρ, for $2g0^2 = e^1 = 1$, and r varying from $r = 0.01$ to $r = 0.9$.

6 Computing ρ and a

The natural parameter of the model are g_0 (gametic standard deviation when there is no homogamy), e (environmental standard deviation), r (correlation between parental phenotypes) and v (environmental correlation between parent and offspring). However we have expressed everything using two extra parameters, a and ρ. We outline here how these two parameters can be computed from the four aforementioned parameters.

We first define a as a function of ρ and the other parameters, excepted v. We have $a^2(1 - r_{A_1A_2}) = 2g_0^2$ where $r_{A_1A_2}$ is

$$r_{A_1A_2} = \frac{r}{\sigma^2}\left(a+\rho e\right)^2 = r\frac{(a+\rho e)^2}{a^2+2\rho ae+e^2},$$

and thus

$$a^2\left(1 - r\frac{(a+\rho e)^2}{a^2+2\rho ae+e^2}\right) = 2g_0^2.$$

This gives an equation in a and ρ that can be solved numerically; it has a unique solution $a(\rho) \in \left[0, \sqrt{\frac{2}{1-r}}g_0\right]$. Then, ρ can be computed by solving

$$v = \rho\frac{a(\rho)^2+2\rho a(\rho)e+e^2+r(\rho a(\rho)+e)^2}{\rho(a(\rho)^2+2\rho a(\rho)e+e^2)+r(\rho a(\rho)+e)(\rho e+a(\rho))}.$$

Note that the right hand term varies from 0 for $\rho = 0$ to 1 for $\rho = 1$. Once the equation is solved for ρ, one naturally takes $a = a(\rho)$. After this, one needs to verify that the value of v is acceptable, that is if the constraint of equation 5 is satisfied.

As a final remark, note that it is easy to check that if (g_0, e, r, v, a, ρ) satisfy the above constraints, then so does $(\kappa g_0, \kappa e, r, v, \kappa a, \rho)$, for any $\kappa > 0$, which corresponds merely to a rescaling of the phenotype.

In figure 1 we show ρ and a as functions of ν, for $2g_0^2 = e^2 = 1$ (corresponding to a phenotype with heritability $h^2 = 0.5$ when $r = \nu = 0$), for several values of r. The point at the end of each line corresponds to the maximal value that ν can take (see section 5.3).

References

[1] Armando Caballero. *Quantitative genetics.* Cambridge University Press, 2020.

[2] James F Crow and Joseph Felsenstein. The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly*, 15(2):85–97, 1968.

[3] Douglas Scott Falconer. *Introduction to Quantitative Genetics.* Oliver & Boyd, London, 1st edition, 1960.

[4] R A Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.

[5] Thomas Nagylaki. The correlation between relatives with assortative mating. *Annals of human genetics*, 42(1):131–137, 1978.

[6] Sewall Wright. Systems of mating. iii. assortative mating based on somatic resemblance. *Genetics*, 6(2):144, 1921.

A Technical lemma

If A, B, X are three random variables, the partial correlation $r_{AB\cdot X}$ is the correlation of A and B conditionnal to X.

Lemma 1 (Partial correlation) *Let (A, B, X) be a Gaussian vector. If A and B are independent conditionnal to X, or equivalently $r_{AB\cdot X} = 0$, then $r_{AB} = r_{AX} r_{BX}$.*

Proof Assume without lost of generality $\mathrm{var}(A) = \mathrm{var}(B) = \mathrm{var}(X) = 1$. The variance-covariance matrix of (A, B, X) is

$$\mathrm{var}(A, B, X) = \begin{pmatrix} 1 & r_{AB} & r_{AX} \\ r_{AB} & 1 & r_{BX} \\ r_{AX} & r_{BX} & 1 \end{pmatrix}.$$

Standard Gaussian theory allows to compute $\mathrm{var}(A, B \mid X)$:

$$\mathrm{var}(A, B \mid X) = \begin{pmatrix} 1 & r_{AB} \\ r_{AB} & 1 \end{pmatrix} - \begin{pmatrix} r_{AX} \\ r_{BX} \end{pmatrix} \begin{pmatrix} r_{AX} & r_{BX} \end{pmatrix} = \begin{pmatrix} 1 - r_{AX}^2 & r_{AB} - r_{AX} r_{BX} \\ r_{AB} - r_{AX} r_{BX} & 1 - r_{BX}^2 \end{pmatrix},$$

implying $r_{AB\cdot X} = r_{AB} - r_{AX} r_{BX}$. □

Lemma 2 *Let (X_1, X_2, Y_1, Y_2) be a gaussian vector. Note $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$. Assume*

$$\mathrm{var}(Y) = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$$

and

$$\mathrm{cov}(X, Y) = \begin{pmatrix} u & v \\ w & w \end{pmatrix}.$$

If X_1, X_2 are independent conditionally to Y, or $\mathrm{cov}(X_1, X_2 \mid Y) = 0$, then

$$\mathrm{cov}(X_1, X_2) = \frac{u + v}{1 + \alpha} w.$$

Proof The variance-covariance matrix of X conditional to Y is

$$\operatorname{var}(X) - \operatorname{cov}(X,Y)\operatorname{var}(Y)^{-1}\operatorname{cov}(Y,X) = \frac{1}{1-\alpha^2}\begin{pmatrix} (1-\alpha)2uv + (u-v)^2 & (1-\alpha)(u+v)w \\ (1-\alpha)(u+v)w & (1-\alpha)2w^2 \end{pmatrix}$$

from which the result is obtained readily. □