Supplementary Material
Twin Research and Human Genetics

Epi2Loc: An R package to investigate two-locus epistatic models

Raymond K. Walters, Charles Laurin, Gitta H. Lubke

S1 Model Design Matrices

Each of the GLM model parameterizations is defined by the design matrix that links the parameters $\boldsymbol{\beta}$ to (a function of) the conditional expected values of the phenotype (Equation 7). The design matrices and parameters for each of the GLM models included in the Epi2Loc package are provided below. Additional details on the motivation and interpretation of each model can be found in their respective sources.

F_2 Model

Adapting the notation of Zeng, Wang, and Zou (2005), the design matrix for the F_2 model (Anderson & Kempthorne, 1954) is defined as

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -\frac{1}{2} & -1 & -\frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & -1 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & -1 & -\frac{1}{2} & -1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & -\frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & -\frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \mu \\ a_a \\ d_a \\ a_b \\ d_b \\ aa_{ab} \\ da_{ab} \\ ad_{ab} \\ dd_{ab} \end{bmatrix}$$

This is a natural extension from coding the effects of a single locus as

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -\frac{1}{2} \\ 1 & 0 & \frac{1}{2} \\ 1 & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}$$

F_∞ Model

Adapting the notation of Zeng et al. (2005), the design matrix for the F_∞ model (Hayman & Mather, 1955) is defined as

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ a_a \\ d_a \\ a_b \\ d_b \\ aa_{ab} \\ da_{ab} \\ ad_{ab} \\ dd_{ab} \end{bmatrix}$$

This is a natural extension of coding the effects of a single locus as

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}$$

G2A Model

The General Two-Allele Model (Zeng et al., 2005) incorporates the marginal allele frequencies for each locus p_a and p_b, respectively. Similarly, denote the major allele frequencies as q_a and q_b, respectively. The resulting model is specified with the following design matrix.

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & -2p_a & -2p_a^2 & -2p_b & -2p_b^2 & 4p_ap_b & 4p_a^2p_b & 4p_ap_b^2 & 4p_a^2p_b^2 \\ 1 & q_a-p_a & 2p_aq_a & -2p_b & -2p_b^2 & 2(p_a-q_a)p_b & -4p_aq_ap_b & 2(p_a-q_a)p_b^2 & -4p_aq_ap_b^2 \\ 1 & 2q_a & -2q_a^2 & -2p_b & -2p_b^2 & -4q_ap_b & 4q_a^2p_b & -4q_ap_b^2 & 4q_a^2p_b^2 \\ 1 & -2p_a & -2p_a^2 & q_b-p_b & 2p_bq_b & 2p_a(p_b-q_b) & 2p_a^2(p_b-q_b) & -4p_ap_bq_b & -4p_a^2p_bq_b \\ 1 & q_a-p_a & 2p_aq_a & q_b-p_b & 2p_bq_b & (p_a-q_a)(p_b-q_b) & 2p_aq_a(q_b-p_b) & 2(q_a-p_a)p_bq_b & 4p_aq_ap_bq_b \\ 1 & 2q_a & -2q_a^2 & q_b-p_b & 2p_bq_b & -2q_a(p_b-q_b) & 2q_a^2(p_b-q_b) & 4q_ap_bq_b & -4q_a^2p_bq_b \\ 1 & -2p_a & -2p_a^2 & 2q_b & -2q_b^2 & -4p_aq_b & -4p_a^2q_b & 4p_aq_b^2 & 4p_a^2q_b^2 \\ 1 & q_a-p_a & 2p_aq_a & 2q_b & -2q_b^2 & -2(p_a-q_a)q_b & 4p_aq_aq_b & 2(p_a-q_a)q_b^2 & -4p_aq_aq_b^2 \\ 1 & 2q_a & -2q_a^2 & 2q_b & -2q_b^2 & 4q_aq_b & -4q_a^2q_b & -4q_aq_b^2 & 4q_b^2q_a^2 \end{bmatrix} \begin{bmatrix} \mu \\ a_a \\ d_a \\ a_b \\ d_b \\ aa_{ab} \\ da_{ab} \\ ad_{ab} \\ dd_{ab} \end{bmatrix}$$

This corresponds to coding the effects at a single locus with allele frequency p as

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -2p & -2p^2 \\ 1 & 1-2p & 2p(1-p) \\ 1 & 2(1-p) & -2(1-p)^2 \end{bmatrix} \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}$$

Unweighted Model

Cheverud and Routman (1995) proposed a model that uses the unweighted mean of the genotype values as the overall mean. The design matrix for this model is given by

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -\frac{1}{3} & -1 & -\frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{9} \\ 1 & 0 & \frac{2}{3} & -1 & -\frac{1}{3} & 0 & -\frac{2}{3} & 0 & -\frac{2}{9} \\ 1 & 1 & -\frac{1}{3} & -1 & -\frac{1}{3} & -1 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{9} \\ 1 & -1 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & -\frac{2}{3} & -\frac{2}{9} \\ 1 & 0 & \frac{2}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 & \frac{4}{9} \\ 1 & 1 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \frac{2}{3} & -\frac{2}{9} \\ 1 & -1 & -\frac{1}{3} & 1 & -\frac{1}{3} & -1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{9} \\ 1 & 0 & \frac{2}{3} & 1 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & -\frac{2}{9} \\ 1 & 1 & -\frac{1}{3} & 1 & -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{9} \end{bmatrix} \begin{bmatrix} \mu \\ a_a \\ d_a \\ a_b \\ d_b \\ aa_{ab} \\ da_{ab} \\ ad_{ab} \\ dd_{ab} \end{bmatrix}$$

where effects for a single locus could be similarly coded as

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -\frac{1}{3} \\ 1 & 0 & \frac{2}{3} \\ 1 & 1 & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}$$

NOIA Model

Alvarez-Castro and Carlborg (2007) propose two forms of the NOIA model. Let p_{AA}, p_{Aa} and p_{aa} be the genotype frequencies for the indicated genotypes at the first locus, and let $\mu_a = p_{Aa} + 2p_{aa}$ be the mean genotype at the first locus. Similarly define p_{BB}, p_{Bb} p_{bb}, and μ_b for the second locus. Then the NOIA Functional model is defined by

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} =$$

$$\begin{bmatrix} 1 & -\mu_a & -p_{Aa} & -\mu_b & -p_{Bb} & \mu_a\mu_b & p_{Aa}\mu_b & \mu_a p_{Bb} & p_{Aa}p_{Bb} \\ 1 & 1-\mu_a & 1-p_{Aa} & -\mu_b & -p_{Bb} & -(1-\mu_a)\mu_b & -(1-p_{Aa})\mu_b & -(1-\mu_a)p_{Bb} & -(1-p_{Aa})p_{Bb} \\ 1 & 2-\mu_a & -p_{Aa} & -\mu_b & -p_{Bb} & -(2-\mu_a)\mu_b & p_{Aa}\mu_b & -(2-\mu_a)p_{Bb} & p_{Aa}p_{Bb} \\ 1 & -\mu_a & -p_{Aa} & 1-\mu_b & 1-p_{Bb} & -\mu_a(1-\mu_b) & -p_{Aa}(1-\mu_b) & -\mu_a(1-p_{Bb}) & -p_{Aa}(1-p_{Bb}) \\ 1 & 1-\mu_a & 1-p_{Aa} & 1-\mu_b & 1-p_{Bb} & (1-\mu_a)(1-\mu_b) & (1-p_{Aa})(1-\mu_b) & (1-\mu_a)(1-p_{Bb}) & (1-p_{Aa})(1-p_{Bb}) \\ 1 & 2-\mu_a & -p_{Aa} & 1-\mu_b & 1-p_{Bb} & (2-\mu_a)(1-\mu_b) & -p_{Aa}(1-\mu_b) & (2-\mu_a)(1-p_{Bb}) & -p_{Aa}(1-p_{Bb}) \\ 1 & -\mu_a & -p_{Aa} & 2-\mu_b & -p_{Bb} & -\mu_a(2-\mu_b) & -p_{Aa}(2-\mu_b) & \mu_a p_{Bb} & p_{Aa}p_{Bb} \\ 1 & 1-\mu_a & 1-p_{Aa} & 2-\mu_b & -p_{Bb} & (1-\mu_a)(2-\mu_b) & (1-p_{Aa})(2-\mu_b) & -(1-\mu_a)p_{Bb} & -(1-p_{Aa})p_{Bb} \\ 1 & 2-\mu_a & -p_{Aa} & 2-\mu_b & -p_{Bb} & (2-\mu_a)(2-\mu_b) & -p_{Aa}(2-\mu_b) & -(2-\mu_a)p_{Bb} & p_{Aa}p_{Bb} \end{bmatrix} \begin{bmatrix} R \\ a_a \\ d_a \\ a_b \\ d_b \\ aa_{ab} \\ da_{ab} \\ ad_{ab} \\ dd_{ab} \end{bmatrix}$$

This is an extension of the following single-locus model

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -\mu_a & -p_{Aa} \\ 1 & 1-\mu_a & 1-p_{Aa} \\ 1 & 2-\mu_a & -p_{Aa} \end{bmatrix} \begin{bmatrix} R \\ a \\ d \end{bmatrix}$$

The NOIA Statistical model is similarly defined by the same expansion of the single-locus model

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & -\mu_a & -\frac{2p_{Aa}p_{aa}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \\ 1 & 1-\mu_a & \frac{4p_{AA}p_{aa}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \\ 1 & 2-\mu_a & -\frac{2p_{AA}p_{Aa}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix}$$

yielding a two-locus model with parameters $\boldsymbol{\beta} = [\mu, \alpha_a, \delta_a, \alpha_b, \delta_b, (\alpha\alpha)_{ab}, (\delta\alpha)_{ab}, (\alpha\delta)_{ab}, (\delta\delta)_{ab}]'$. The full design matrix is omitted here due to space limitations.

Genotype Model

Finally, the genotype model described by Cordell (2009, Supplementary Box S1), is defined by

$$\begin{bmatrix} u_{AABB} \\ u_{AaBB} \\ u_{aaBB} \\ u_{AABb} \\ u_{AaBb} \\ u_{aaBb} \\ u_{AAbb} \\ u_{Aabb} \\ u_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma_1 \\ \gamma_2 \\ \beta_1 \\ \beta_2 \\ i_{11} \\ i_{12} \\ i_{21} \\ i_{22} \end{bmatrix}$$

As with the other models, the interaction parameters for the genotype model correspond to cross-products of a single-locus model.

$$\begin{bmatrix} u_{AA} \\ u_{Aa} \\ u_{aa} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma_1 \\ \gamma_2 \end{bmatrix}$$

S2 NOIA Model Variance Components

The orthogonal parameters of the NOIA Statistical model (Alvarez-Castro & Carlborg, 2007) correspond to the orthogonal variance components for partitioning the total variance explained by the two loci V_g (Equations 4, 6). This enables direct conversion between the GLM model parameters and variance components. The relationship between the two models is detailed below.

$$V_{A_a} = \alpha_a^2 \left[p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2\right]$$

$$V_{D_a} = \delta_a^2 \left[\frac{4p_{AA}p_{Aa}p_{aa}}{p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2}\right]$$

$$V_{A_b} = \alpha_b^2 \left[p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2\right]$$

$$V_{D_b} = \delta_b^2 \left[\frac{4p_{BB}p_{Bb}p_{bb}}{p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2}\right]$$

$$V_{AA} = (\alpha\alpha)_{ab}^2 \left[p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2\right] \left[p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2\right]$$

$$V_{DA} = (\delta\alpha)_{ab}^2 \left[\frac{4p_{AA}p_{Aa}p_{aa}}{p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2}\right] \left[p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2\right]$$

$$V_{AD} = (\alpha\delta)_{ab}^2 \left[p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2\right] \left[\frac{4p_{BB}p_{Bb}p_{bb}}{p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2}\right]$$

$$V_{DD} = (\delta\delta)_{ab}^2 \left[\frac{4p_{AA}p_{Aa}p_{aa}}{p_{Aa}(1-p_{Aa}) + 4p_{aa}(1-p_{Aa}) - 4p_{aa}^2}\right] \left[\frac{4p_{BB}p_{Bb}p_{bb}}{p_{Bb}(1-p_{Bb}) + 4p_{bb}(1-p_{Bb}) - 4p_{bb}^2}\right]$$

References

Alvarez-Castro, J. M., & Carlborg, O. (2007). A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics*, *176*, 1151–1167.

Anderson, V. L., & Kempthorne, O. (1954). A model for the study of quantitative inheritance. *Genetics*, *39*, 883–898.

Cheverud, J. M., & Routman, E. J. (1995). Epistasis and its contribution to genetic variance components. *Genetics*, *139*, 1455–1461.

Cordell, H. J. (2009). Detecting genegene interactions that underlie human diseases. *Nature Reviews Genetics*, *10*, 392–404.

Hayman, B., & Mather, K. (1955). The description of genic interactions in continuous variation. *Biometrics*, *11*(1), 69–82.

Zeng, Z.-B., Wang, T., & Zou, W. (2005). Modeling quantitative trait loci and interpretation of models. *Genetics*, *169*, 1711–1725.