File S2: Formalism with random individuals and allele effects

In the RI formalism, allele effects are assumed to be deterministic whereas individuals (both genotype and allele ancestry) are assumed to be random. In the RAE formalism, the reverse situation is considered. We define here a new formalism in which both individuals and allele effects are considered random. By deriving the covariance between genetic values, we evaluate whether this new formalism helps to define a new genomic prediction model, as illustrated here on GBLUP.

The GBLUP-RI covariance is a covariance over genotypes (Cov_W) conditional to trait parameters (i.e. allele effects β_m^1 and β_m^0), population parameters (i.e. allele frequencies f_m) and a parameter indicating the level of resemblance between individuals (i.e. kinship α_{ij}):

$$\mathrm{Cov}_W(G_i, G_j | \alpha_{ij}, f_m, \beta_m^1, \beta_m^0) = \alpha_{ij} \sum_{m=1}^{M} f_m(1 - f_m) \left(\beta_m^1 - \beta_m^0\right)$$

It should be understood as the covariance for a given trait between the genetic values of random individuals having a given kinship and belonging to the same population.

The GBLUP-RAE covariance is a covariance over allele effects (Cov_β) conditional to observed data (i.e. genotypes w_{im} and w_{jm}) and trait parameters (i.e. the distribution of allele effects $\mathcal{N}(0, \sigma_\beta^2)$):

$$\mathrm{Cov}_\beta(G_i, G_j | w_{im}, w_{jm}, \sigma_\beta^2) = \sum_{m=1}^{M} \left((1 - w_{im})(1 - w_{jm}) + w_{im}w_{jm}\right) \sigma_\beta^2$$

It should be understood as the covariance between the genetic values of given individuals for a random trait with a given distribution of allele effects.

Using the new formalism where both genotypes and allele effects are random, we can apply the law of the total covariance based on the GBLUP-RI expected value and covariance:

$$\begin{aligned} \mathrm{Cov}(G_i, G_j | \alpha_{ij}, f_m, \sigma_\beta^2) = &\mathrm{Cov}_\beta \left[\mathrm{E}_W\left(G_i | f_m, \beta_m^1, \beta_m^0\right), \mathrm{E}_W\left(G_j | f_m, \beta_m^1, \beta_m^0\right) | \sigma_\beta^2\right] \\ &+ \mathrm{E}_\beta \left[\mathrm{Cov}_W(G_i, G_j | \alpha_{ij}, f_m, \beta_m^1, \beta_m^0) | \sigma_\beta^2\right] \end{aligned}$$

It is also possible to apply the law of the total covariance based on the GBLUP-RAE expected value and covariance:

$$\begin{aligned} \mathrm{Cov}(G_i, G_j | \alpha_{ij}, f_m, \sigma_\beta^2) = &\mathrm{Cov}_W \left[\mathrm{E}_\beta\left(G_i | w_{im}, \sigma_\beta^2\right), \mathrm{E}_\beta\left(G_j | w_{jm}, \sigma_\beta^2\right) | \alpha_{ij}, f_m\right] \\ &+ \mathrm{E}_W \left[\mathrm{Cov}_\beta(G_i, G_j | w_{im}, w_{jm}, \sigma_\beta^2) | \alpha_{ij}, f_m\right] \end{aligned}$$

Both approaches lead to the same expression:

$$\mathrm{Cov}(G_i, G_j | \alpha_{ij}, f_m, \sigma_\beta^2) = \sum_{m=1}^{M} \left((1 - f_m)^2 + f_m^2\right) \sigma_\beta^2 + \alpha_{ij} \sum_{m=1}^{M} 2f_m(1 - f_m)\sigma_\beta^2$$

The resulting covariance should be interpreted as the covariance between the genetic values of random individuals having a given kinship and belonging to the same population for a random trait with a given distribution of allele effects. As the resulting covariance is expressed as the sum of a first term that does not depend on i and j and a second term that is proportional to α_{ij} (same as in GBLUP-RI), this new formalism leads to the same genomic prediction model as GBLUP-RI.