Variance and Covariance of Actual Relationships between Relatives at One Locus


Luis Alberto Garcia-Cortés, Andrés Legarra, Claude Chevalet, Miguel Angel Toro,

EAAP 2013
Why

Realized relationships = expected relationships + deviations

\[ G = A + E \]

Estimated with markers (Van Raden 2008 and many others)

Estimated with pedigree (Emik & Terrill 1949 and many others)

?????
Some sibs are more equal than others!!
(Invented) Full-sibs example

<table>
<thead>
<tr>
<th>$A_{ij}$</th>
<th>$G_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.48</td>
</tr>
<tr>
<td>0.5</td>
<td>0.60</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.70</td>
</tr>
<tr>
<td>0.5</td>
<td>0.44</td>
</tr>
<tr>
<td>0.5</td>
<td>0.53</td>
</tr>
</tbody>
</table>

$Var(G_{ij}) = Var(G_{ij} - A_{ij}) = 0.023$

Can we compute this variance from pedigree before seeing the markers?
Why

• Increased accuracy of genomic selection relies on using a more accurate $G$ instead of $A$
  - But we don’t know a priori how much will they differ

• Small errors in $A$ cumulate with generations !!
  - How many generations should one use?

• For both questions we need to know the extent of deviations in $E = G - A$
Variation in actual relationship as a consequence of Mendelian sampling and linkage

W.G. Hill¹ and B.S. Weir²

• Only consider two-individual coancestries and not a full pedigree
• Not a general formula but a collection of them
• They consider linkage, we consider one locus
• We provide a general formula for all pairs of individuals in a pedigree
What

Realized relationships = expected relationships + deviations
\[ G = A + E \]

What is \( V \var{E} \)?
Let \( E = (e_1, e_2, \ldots, e_n) \)

What is \( \text{Cov} \left( \begin{array}{c} e_1 \\ e_2 \\ \vdots \\ e_n \end{array} \right) \)?

Coancestry \( \Phi \)
\[ V \var{E} = 4V \var{\Phi} \]
\[ \Delta_i = \text{Probability of identity state } S_i \]

Father-son

\[ \Delta_8 = 1 \]

Full sibs

\[ \Delta_7 = \frac{1}{4} \]

\[ \Delta_8 = \frac{1}{2} \]

\[ \Delta_9 = \frac{1}{4} \]

Coancestry \((= 0.5a_{ij})\) of \(i\) and \(j\)

\[
\phi_{ij} = \Delta_1 + \frac{1}{2} \left( \Delta_3 + \Delta_5 + \Delta_7 \right) + \lambda_1 \Delta_8
\]

Fig. 1. Condensed identity states.
\[ E(i) = \Delta_i \]
\[ \text{Var}(i) = \Delta_i (1 - \Delta_i) \]
\[ \text{Cor}(\epsilon_i') = -\Delta_i \Delta_j \]

\[ \phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \gamma \Delta_8 \]

\[ \phi_{ij} = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_q \end{pmatrix} \times \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_q \end{pmatrix} \]

\[ \text{Var}(\phi_{ij}) = \text{Var}(W' \Delta) \]
\[ = W' \text{Var}(\Delta) W \]
\[ = \sum_{i} w_i^2 \Delta_i (1 - \Delta_i) - 2 \sum_{i} \sum_{j} w_i w_j \Delta_i \Delta_j \]

\[ \text{Var}(\phi_{ij}) = \sum_{i} w_i^2 \Delta_i - \left(\sum_{i} w_i \Delta_i\right)^2 \]
\[ = \phi_{ij}, j' \quad \text{regular coancestry} \]

Karigl (1981) generalized relationship
The covariance is a simple extension:
\[
\text{Cov}(\Phi_{AB}, \Phi_{CD}) = \phi_{AB,CD} - \phi_{AB}\phi_{CD}
\]

This means that the \( \Phi_{CD} \) is not independent of \( \Phi_{AB} \),
example:

C and D cannot share alleles if A and B do not
Generalized relationships

Karigl’s 1981 generalized relationships

\[ \phi_{AB,CD} \]

is the probability that

two genes drawn from A and B are IBD
and at the same time
two genes drawn from C and D are IBD

Can be computed from pedigree
It will depend on the paths linking all four individuals
Inbreeding of Dash 2nd

\[
F_D = 0.4297
\]

\[
\phi_{CS} = 0.4297
\]

\[
V(F_D) = 0.2451
\]

\[
V(\phi_{CS}) = 0.0922
\]

Coancestry of parents of Dash 2nd

\[F_D = \phi_{CS}, \text{ yet the variances are different}!!\]
Extension to several loci

- If there were infinite loci, these variances are divided by $\infty$ and become 0
- Genome is not infinite, is composed of repetitive “chunks” due to linkage and finite number of animals
- Divide by « Me = equivalent number of loci » (Goddard 2009 and many others)

$$Cov(\Phi_{AB}, \Phi_{CD}) = \frac{1}{Me} (\Phi_{AB,CD} - \Phi_{AB} \Phi_{CD})$$
Does it work?

- H. Wang, I. Misztal et al., in prep, chicken data, >4000 animals
- \( E = G - A \)
- empirical \( \text{var}(E) \) vs. predictions based on pedigree using \( \text{cov}(\Phi_{AB}, \Phi_{CD}) = \phi_{AB,CD} - \phi_{AB}\phi_{CD} \)
Var(G-A) vs our theory

- Pretty good fit
- From here, number of effective loci ~70
Conclusions

• $Var(E = G - A)$ can be computed from pedigree and a populational parameter ($Me$)

• Starting point for more developments
Acknowledgments

• project CGL2009-1327-C02-02 of the Ministerio de Ciencia e Innovacion (Spain)
• ANR project Rules & Tools (France)
• action X-Gen of the SelGen metaprogram (INRA)
• partially supported by the platform bioinformatics Toulouse Midi-Pyrenees

• LAGC made the drawings