UNIL | Université de Lausanne
Département d'écologie et évolution

DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

A consistent estimator of kinship for admixed populations, applied to heritability estimation.

Jérôme Goudet (jerome.goudet@unil.ch) and Bruce Weir (bsweir@uw.edu)

Quantitative Trait Variance

For GRM $\boldsymbol{G}$ with elements $(1+F_i)$ on the diagonal, and elements $2\theta_{ij}$ off the diagonal, the $n \times 1$ vector $\boldsymbol{Y}$ of trait values has variance

$$\mathrm{Var}(\boldsymbol{Y}) = \boldsymbol{G}\sigma_A^2 + \boldsymbol{I}\sigma_e^2$$

The trace of this matrix is

$$\mathrm{tr}(\boldsymbol{G}) = \sum_{i=1}^{n} G_{ii} = n(1+F_W)$$

to define the average inbreeding F_W for the sample, and the sum of the off-diagonal elements is

$$\Sigma_{\boldsymbol{G}} = \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} G_{ij} = 2n(n-1)\theta_S$$

to define the average kinship θ_S for the sample.

Speed et al.

Speed et al. calculated two variances, $\hat{V}_Y$ for the sample variance of trait values and $\hat{V}_R$ for the residual variance once the genotypic effects have been fitted, to estimate heritability:

$$\widehat{h^2} = \frac{\hat{V}_Y - \hat{V}_R}{\hat{V}_Y}$$

If F_W, θ_S are known, we find that

$$\mathcal{E}(\widehat{h^2}) = \frac{(1+F_W-2\theta_S)\sigma_A^2}{(1+F_W-2\theta_S)\sigma_A^2+\sigma_e^2}$$

to emphasize the role of both inbreeding and kinship.

Use of Allele-sharing GRM

Weir & Goudet, 2017, estimate half the GRM by $\boldsymbol{K}_{as}$:

$$\boldsymbol{K}_{\mathrm{as}_{\mathrm{ij}}} = \frac{\tilde{M}_{ij} - \tilde{M}_S}{1-\tilde{M}_S}$$

$\tilde{M}_{ij}$ is allele-sharing for individuals i, j, with mean $\tilde{M}_S$ over $i \neq j$. In this case $\Sigma_{\boldsymbol{K}_{\mathrm{as}}}$ is 0 and $\mathcal{E}[\mathrm{tr}(\boldsymbol{K}_{as})]$ is $n(1+f_W)/2$, where $f_W = (F_W - \theta_S)/(1-\theta_S)$ is the within-population inbreeding coefficient (i.e. F_{IS}). Therefore

$$\mathcal{E}(\widehat{h^2}) = \frac{(1+f_W)\sigma_A^2}{(1+f_W)\sigma_A^2+\sigma_e^2}$$

This replaces F_W (i.e. F_{IT}) in the classical result with f_W, reflecting that it is f_W and not F_W that can be estimated with data from a single population.

Use of Standard GRM

When half the GRM is estimated by $\boldsymbol{K}_{\mathrm{Std}}$ with elements

$$\frac{\sum_l (X_{il}-2\tilde{p}_l)(X_{jl}-2\tilde{p}_l)}{\sum_l 4\tilde{p}_l(1-\tilde{p}_l)}$$

we find for large n, the same estimate of heritability, even though $\boldsymbol{K}_{\mathrm{Std}} \neq \boldsymbol{K}_{\mathrm{as}}$.

GCTA (Yang et al, 2011) modifies the diagonal elements of the GRM to

$$\frac{1}{L}\sum_{l=1}^{L}\frac{[X_{il}^2 - 2(1+\tilde{p}_l)X_{il} + 2\tilde{p}_l]}{4\tilde{p}_l(1-\tilde{p}_l)}$$

If the average over SNPs of ratios is changed to the ratio of averages, this matrix also gives the same heritability estimate as does $\boldsymbol{K}_{\mathrm{as}}$.

Relationship Between GRMs

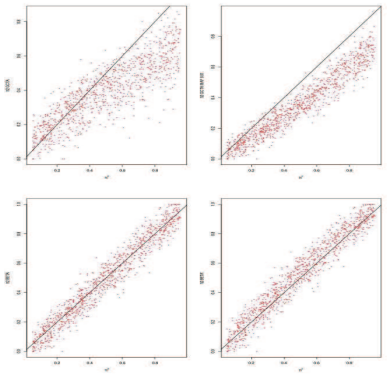
For large sample sizes, the standard matrix is the double-centered allele-sharing one:

$$\boldsymbol{K}_{\mathrm{Std}} = (\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{K}_{\mathrm{as}}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})$$

where $\boldsymbol{J}$ is an $n \times n$ matrix with every element equal to 1.

We also find that, for large n, $\mathrm{tr}(\boldsymbol{K}_{\mathrm{as}})$, $\mathrm{tr}(\boldsymbol{K}_{\mathrm{Std}})$, $\mathrm{tr}(\boldsymbol{K}_{\mathrm{GCTA}})$ are all the same, but only $\boldsymbol{K}_{\mathrm{as}}$ gives inbreeding and kinship estimates that rank individuals consistently across different reference sample sets.

Numerical Results



Results for 1,000 simulated traits using 1000 Genomes data. 10,000 causal loci were drawn randomly from chrs 1 and 2. X axis shows true heritabilities, Y axis shows estimates,

Top row uses $\boldsymbol{K}_{\mathrm{GCTA}}$ with average of ratios for combining SNPs. Bottom row uses $\boldsymbol{K}_{\mathrm{as}}$ as shown above.

Left column uses all SNPs, right column uses only SNPs with MAF>0.01.

References

Speed et al. 2012 AJHG 91:1011
Weir, Goudet. 2017 Genetics 206:2085
Yang et al. 2011 AJHG 88:76

Support

Swiss NSF 31003A-138180,IZKOZ3-157867, US NIH GM075091.