HERITABILITY ESTIMATION

Heritability

For an additive trait, heritability in a HWE population is

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2}$$

Estimation of h^2 therefore requires estimation of σ_A^2 and σ_E^2. There are likelihood-based methods for doing that, assuming the trait values are normally distributed.

These notes follow a discussion given by Speed D, et al. 2012. Am J Hum Genet 91:1011

Speed et al. 2012

Instead of having replicates of the history of a single individual, use the trait values for a sample of individuals in some population. Speed et al. used $\widehat{V}_T$ for the sample variance of trait values:

$$\widehat{V}_T \;=\; \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$$

and $\widehat{V}_R$ for the residual variance once the genotypic effects have been fitted:

$$\widehat{V}_R \;=\; \frac{1}{n-1}\sum_i (E_i - \bar{E})^2$$

As an estimate of heritability, Speed et al. combined these two sample variances

$$\widehat{h^2} \;=\; \frac{\widehat{V}_T - \widehat{V}_R}{\widehat{V}_T}$$

Speed et al. 2012

It can be shown that

$$\mathcal{E}(\hat{V}_T) = \frac{1}{n}\left[\text{tr}(\boldsymbol{G}) - \frac{1}{n-1}\Sigma_{\boldsymbol{G}}\right]\sigma_A^2 + \sigma_E^2$$

$$\mathcal{E}(\hat{V}_R) = \sigma_E^2$$

so that

$$\mathcal{E}(\widehat{h^2}) = \frac{\frac{1}{n}\left[\text{tr}(\boldsymbol{G}) - \frac{1}{n-1}\Sigma_{\boldsymbol{G}}\right]\sigma_A^2}{\frac{1}{n}\left[\text{tr}(\boldsymbol{G}) - \frac{1}{n-1}\Sigma_{\boldsymbol{G}}\right]\sigma_A^2 + \sigma_E^2}$$

and this has a parametric value of

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

Expectation of $\widehat{h^2}$

In the case of no identity by descent within or between individuals, $F_W = \theta_S = 0$,

$$\mathcal{E}(\widehat{h^2}) \;=\; \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} = h^2$$

In general, however, the expectation of $\widehat{h}^2$ is

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

where $f = (F_W - \theta_S)/(1 - \theta_S)$. We are not aware that this expression has been presented before, although

$$\mathcal{E}(\widehat{h^2}) \;=\; \frac{(1 + F)\sigma_A^2}{(1 + F)\sigma_A^2 + \sigma_E^2}$$

is often given as though the study population has inbreeding but no kinship: $F = f, \theta_S = 0$. It does not seem likely that a natural population could consist of inbred but unrelated individuals.

Expectation of $\widehat{h^2}$

For a population in Hardy-Weinberg equilibrium, $F_W = \theta_S, f = 0$:

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

and then $\widehat{h^2}$ will be close to unbiased if θ_S is low.

Use of Estimated GRM

The Speed et al. estimate uses two sample variances, and does not make explicit use of the GRM. Likelihood-based methods do use the GRM. As the parametric values F_i, θ_{ij} are not generally known, $\boldsymbol{G}$ is replaced by a matrix $\boldsymbol{K}$ of estimates. What is the resulting heritability estimate then estimating?

Use of Allele-sharing GRM

Can estimate half the GRM with $\hat{K}_{as}$ having elements $\{\hat{\beta}_{ij}\}$. If $\tilde{M}_{ij}$ is the allelic matching proportion, averaged over SNPs, for individuals i and j including $i = j$, the β estimates are

$$\hat{\beta}_{ij} \;=\; \frac{\tilde{M}_{ij} - \tilde{M}_S}{1 - \tilde{M}_S}$$

where $\tilde{M}_{ij} = \sum_{l=1}^{L}[1 + (X_{il} - 1)(X_{jl} - 1)]/(2L)$ for allelic dosages X_{il} and $\tilde{M}_S = \sum_{i \neq j} \tilde{M}_{ij}/[n(n-1)]$. These estimates have expected values

$$\mathcal{E}(\hat{\beta}_{ij}) \;=\; \begin{cases} \frac{\frac{1}{2}(1+F_i) - \theta_S}{1 - \theta_S} & i = j \\[2ex] \frac{\theta_{ij} - \theta_S}{1 - \theta_S} & i \neq j \end{cases}$$

Use of Allele-sharing GRM

As $\Sigma_{\hat{K}_{as}} = 0$ by construction, the expectation of the estimated heritability is

$$\mathcal{E}(\hat{h}^2) = \frac{\mathcal{E}[\frac{2}{n}\text{tr}(\hat{K}_{as})\hat{\sigma}_A^2]}{\mathcal{E}[\frac{2}{n}\text{tr}(\hat{K}_{as})\hat{\sigma}_A^2 + \hat{\sigma}_E^2]}$$

From the expected values of $\hat{\beta}_{ij}$, $\mathcal{E}[\text{tr}(\hat{K}_{as})] = n(1+f)/2$ is assumed known and replaces $\hat{K}_{as}$, leading to

$$\mathcal{E}(\hat{h}^2) = \frac{(1+f)\sigma_A^2}{(1+f)\sigma_A^2 + \sigma_E^2}$$

This replaces F in the classical result with f, reflecting that is f and not F that can be estimated with data from a single population.

Use of GCTA GRM

Can also estimate half the GRM with $\hat{K}_c(0)$ having elements $\{\hat{k}_{ij}\}$:

$$\hat{k}_{ij} = \frac{\sum_l (X_{il} - 2\tilde{p}_l)(X_{jl} - 2\tilde{p}_l)}{\sum_l 4\tilde{p}_l(1 - \tilde{p}_l)}$$

Now all the elements of the GRM sum to zero by construction. In other words $\mathrm{tr}[\hat{K}_c(0)] + \Sigma_{\hat{K}_c(0)} = 0$ and the estimated heritability is

$$\hat{h}^2 = \frac{\frac{2}{n-1}\mathrm{tr}[\hat{K}_c(0)]\hat{\sigma}_A^2}{\frac{2}{n-1}\mathrm{tr}[\hat{K}_c(0)]\hat{\sigma}_A^2 + \hat{\sigma}_E^2}$$

Use of GCTA GRM

Since

$$\mathcal{E}(\widehat{k}_{ii}) = \frac{1}{2}\left(1 + \frac{F_i - 2\psi_i + \theta_S}{1 - \theta_S}\right) \ , \ \psi_i = \frac{1}{n-1}\sum_{j \neq i}\theta_{ij} \ , \ \theta_S = \frac{1}{n}\sum_i \psi_i$$

$$\mathcal{E}[\text{tr}[\widehat{K}_c(0)] \ = \ \frac{n}{2}(1 + f)$$

and, regarding this as a constant, the expected value of the estimated heritability is

$$\mathcal{E}(\widehat{h}^2) \ = \ \frac{\frac{n}{n-1}(1 + f)\sigma_A^2}{\frac{n}{n-1}(1 + f)\sigma_A^2 + \sigma_E^2}$$

$$\approx \ \frac{(1 + f)\sigma_A^2}{(1 + f)\sigma_A^2 + \sigma_E^2}$$

as for the β estimate.

Very different GRMs give the same estimates of heritability.