

Heritability

Nuno Carvalho

April 16, 2025

1 SNP Heritability

These derivations are based on the Methods of [Yang et al., 2010].

1.1 Phenotype model

We can define a quantitative phenotype y as:

$$y = Z\beta + \epsilon$$

Where:

- y : phenotypes. $N \times 1$ vector. Centered so that $E[y] = 0$.
 - N : number of samples.
- Z : normalized genotypes for causal SNPs. $N \times M$ matrix.
 - M : number of causal SNPs.
 - Normalized according to $Z_i = \frac{Z'_i - 2f_i}{\sqrt{2f_i(1-f_i)}}$.
 - Z' : allele dosages, taking on values of 0, 1, 2.
 - f_i : true population allele frequency for SNP i .
 - Such that for each row (SNP), $E[Z_i] = 0$ and $Var[Z_i] = 1$.
- β : per-normalized-genotype causal effects. $M \times 1$ vector.
 - Assume infinitesimal model.
 - Drawn from $\beta \sim \mathcal{N}(0, I\sigma_\beta^2)$.
 - I : $M \times M$ identity matrix.
 - σ_β^2 : variance of causal effects.
- ϵ : residual effects (i.e. error or noise term). $N \times 1$ vector.
 - Drawn from $\epsilon \sim \mathcal{N}(0, I\sigma_\epsilon^2)$.
 - I : $N \times N$ identity matrix.
 - σ_ϵ^2 : residual variance.

We assume Z , β , and ϵ are all independent from each other. We can define the genetic effects as a single term, $g = Z\beta$, meaning that:

$$y = g + \epsilon \quad \text{where} \quad g \sim \mathcal{N}(0, I\sigma_g^2) \quad \text{where} \quad \sigma_g^2 = M\sigma_\beta^2$$

We interpret σ_g^2 as variance of total additive genetic effects on the phenotype.

1.2 Variance of the phenotype

By making use of the independence between terms, we can define the variance-covariance matrix of y as:

$$\begin{aligned} Var[y] &= Var[Z\beta + \epsilon] \\ &= Var[Z]Var[\beta] + Var[\epsilon] \\ &= (ZZ^\top)\sigma_\beta^2 + I\sigma_\epsilon^2 \\ &= (ZZ^\top)\frac{\sigma_g^2}{M} + I\sigma_\epsilon^2 \\ &= G\sigma_g^2 + I\sigma_\epsilon^2 \end{aligned}$$

Where we define $G = \frac{ZZ^\top}{M}$ as the $N \times N$ genetic relationship matrix (GRM) between individuals. The G_{ii} element is the variance of individual i 's normalized genotype vector, while the G_{ij} element is the covariance of individuals i and j 's normalized genotype vectors.

Narrow-sense heritability is defined as the proportion of phenotypic variance, σ_P^2 , explained by additive genetic effects:

$$h^2 = \frac{\sigma_g^2}{\sigma_P^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2}$$