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:

$$\mathbf{y} = \mathbf{X}_c\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Where:

- $\mathbf{y}$: phenotypes. $N \times 1$ vector. Centered so that $\mathrm{E}[\mathbf{y}] = 0$.
  - N: number of samples.
- $\mathbf{X}_c$: normalized genotypes for causal variants. $N \times M_c$ matrix.
  - M_c: number of causal variants.
  - Normalized according to $\mathbf{X}_{c,i} = \frac{\mathbf{X}'_{c,i} - 2f_i}{\sqrt{2f_i(1-f_i)}}$.
    - $\mathbf{X}'_c$: allele dosages, taking on values of 0, 1, 2.
    - f_i: true population allele frequency for variant i.
    - Such that for each row (variant), $\mathrm{E}[\mathbf{X}_{c,i}] = 0$ and $\mathrm{Var}[\mathbf{X}_{c,i}] = 1$.
- $\boldsymbol{\beta}$: per-normalized-genotype causal effects. $M_c \times 1$ vector.
  - Assume infinitesimal model.
  - Drawn from $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2_\beta)$.
    - $\mathbf{I}$: $M_c \times M_c$ identity matrix.
    - σ^2_β: variance of causal effects.
- $\boldsymbol{\epsilon}$: residual effects (i.e. error or noise term). $N \times 1$ vector.
  - Drawn from $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}\sigma^2_\epsilon)$.
    - $\mathbf{I}$: $N \times N$ identity matrix.
    - σ^2_ϵ: residual variance.

We assume $\mathbf{X}_c$, $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ are all independent from each other. We can define the genetic effects as a single term, $\mathbf{g} = \mathbf{X}_c\boldsymbol{\beta}$, meaning that:

$$\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon} \quad \text{where} \quad \mathbf{g} \sim \mathcal{N}(0, I\sigma^2_g) \quad \text{where} \quad \sigma^2_g = M_c\sigma^2_\beta$$

We interpret σ^2_g 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 $\mathbf{y}$ as:

$$\begin{aligned}
\text{Var}[\mathbf{y}] &= \text{Var}[\mathbf{X}_c\boldsymbol{\beta} + \boldsymbol{\epsilon}] \\
&= \text{Var}[\mathbf{X}_c]\text{Var}[\boldsymbol{\beta}] + \text{Var}[\boldsymbol{\epsilon}] \\
&= (\mathbf{X}_c\mathbf{X}_c^\intercal)\sigma_g^2 + \mathbf{I}\sigma_\epsilon^2 \\
&= (\mathbf{X}_c\mathbf{X}_c^\intercal)\frac{\sigma_g^2}{M_c} + \mathbf{I}\sigma_\epsilon^2 \\
&= \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_\epsilon^2
\end{aligned}$$

Where we define $\mathbf{G} = \frac{\mathbf{X}_c\mathbf{X}_c^\intercal}{M_c}$ 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:

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