Contrasting the genetic architecture of 30 complex traits from summary association data (additional notes)

Huwenbo Shi

What does HESS estimate?

Linear model for a genome partitioned into m (not necessarily independent) windows says

$$y = \mathbf{x}_1^T \boldsymbol{\beta}_1 + \cdots \mathbf{x}_m^T \boldsymbol{\beta}_m + \epsilon, \tag{1}$$

where $\boldsymbol{\beta}_i$ is the true effect size vector for the i-th loci. Therefore,

$$\begin{aligned} \mathrm{Var}[y] &= \sum_{i=1}^{m} \mathrm{Var}[\mathbf{x}_i^T \boldsymbol{\beta}_i] + \sum_{i=1}^{m} \sum_{j=1}^{m} \mathrm{Cov}[\mathbf{x}_i^T \boldsymbol{\beta}_i, \mathbf{x}_j^T \boldsymbol{\beta}_j] + \epsilon \\ &= \sum_{i=1}^{m} \boldsymbol{\beta}_i^T \mathbf{V}_i \boldsymbol{\beta}_i + \sum_{i=1}^{m} \sum_{j=1}^{m} \boldsymbol{\beta}_i^T \mathbf{C}_{ij} \boldsymbol{\beta}_j + \epsilon, \end{aligned} \tag{2}$$

where $\mathbf{V}_i$ is the LD matrix for i-th locus, and $\mathbf{C}_{ij}$ is the covariance matrix between SNPs in window i and j. What is being estimated by HESS is the term $\boldsymbol{\beta}_i^T \mathbf{V}_i \boldsymbol{\beta}_i$.

In the special case when SNPs in window i and j are independent, i.e. $\mathbf{C}_{ij} = \mathbf{0}$, summing over the HESS estimates across all loci ($\sum_{i=1}^{m} \hat{h}^2_{g,local,i}$) gives an unbiased estimates of total heritability. However, when $\mathbf{C}_{ij} \neq \mathbf{0}$, the simple summation will give a biased estimates of total heritability, with bias equal to $\sum_{i=1}^{m} \sum_{j=1}^{m} \boldsymbol{\beta}_i^T \mathbf{C}_{ij} \boldsymbol{\beta}_j$. The next section derives an unbiased estimator for $\boldsymbol{\beta}_i^T \mathbf{C}_{ij} \boldsymbol{\beta}_j$.

An (approximately) unbiased estimator for $\boldsymbol{\beta}_i^T \mathbf{C}_{ij} \boldsymbol{\beta}_j$

Let $\hat{\boldsymbol{\beta}}_i = \frac{1}{n} \mathbf{X}_i^T \mathbf{y}$ and $\hat{\boldsymbol{\beta}}_j = \frac{1}{n} \mathbf{X}_j^T \mathbf{y}$ be the effect size vectors estimated through GWAS. It's well known that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i &\sim N\left(\mathbf{V}_i \boldsymbol{\beta}_i, \frac{\sigma_e^2}{n} \mathbf{V}_i\right) \\ \hat{\boldsymbol{\beta}}_j &\sim N\left(\mathbf{V}_j \boldsymbol{\beta}_j, \frac{\sigma_e^2}{n} \mathbf{V}_j\right) \\ \begin{bmatrix} \hat{\boldsymbol{\beta}}_i \\ \hat{\boldsymbol{\beta}}_j \end{bmatrix} &\sim N\left(\begin{bmatrix} \mathbf{V}_i \boldsymbol{\beta}_i \\ \mathbf{V}_j \boldsymbol{\beta}_j \end{bmatrix}, \frac{\sigma_e^2}{n} \begin{bmatrix} \mathbf{V}_i & \mathbf{C}_{ij} \\ \mathbf{C}_{ij}^T & \mathbf{V}_j \end{bmatrix} \right) \end{aligned} . \tag{3}$$

Let $\hat{\boldsymbol{\alpha}}_i = \mathbf{V}_i^{-1}\hat{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\alpha}}_j = \mathbf{V}_j^{-1}\hat{\boldsymbol{\beta}}_j$, then

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_i &\sim N\left(\boldsymbol{\beta}_i, \frac{\sigma_e^2}{n}\mathbf{V}_i^{-1}\right)\\ \hat{\boldsymbol{\alpha}}_j &\sim N\left(\boldsymbol{\beta}_j, \frac{\sigma_e^2}{n}\mathbf{V}_j^{-1}\right).\end{aligned} \tag{4}$$

And

$$\begin{aligned}\mathrm{Cov}[\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\alpha}}_j] &= \mathrm{Cov}[\mathbf{V}_i^{-1}\hat{\boldsymbol{\beta}}_i, \mathbf{V}_j^{-1}\hat{\boldsymbol{\beta}}_j]\\ &= \mathrm{E}[(\mathbf{V}_i^{-1}\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)(\mathbf{V}_j^{-1}\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)^T]\\ &= \mathrm{E}[\mathbf{V}_i^{-1}\hat{\boldsymbol{\beta}}_i\hat{\boldsymbol{\beta}}_j^T\mathbf{V}_j^{-1} - \mathbf{V}_i^{-1}\hat{\boldsymbol{\beta}}_i\boldsymbol{\beta}_j^T - \boldsymbol{\beta}_i\hat{\boldsymbol{\beta}}_j^T\mathbf{V}_j^{-1} + \boldsymbol{\beta}_i\boldsymbol{\beta}_j^T]\\ &= \frac{\sigma_e^2}{n}\mathbf{V}_i^{-1}\mathbf{C}_{ij}\mathbf{V}_j^{-1} - 2\boldsymbol{\beta}_i\boldsymbol{\beta}_j^T + \boldsymbol{\beta}_i\boldsymbol{\beta}_j^T\\ &= \frac{\sigma_e^2}{n}\mathbf{V}_i^{-1}\mathbf{C}_{ij}\mathbf{V}_j^{-1} - \boldsymbol{\beta}_i\boldsymbol{\beta}_j^T.\end{aligned} \tag{5}$$

If we assume $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$ are sparse, i.e. not many causal SNPs, then

$$\mathrm{Cov}[\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\alpha}}_j] \approx \frac{\sigma_e^2}{n}\mathbf{V}_i^{-1}\mathbf{C}_{ij}\mathbf{V}_j^{-1}. \tag{6}$$

An initial estimator for $\boldsymbol{\beta}_i^T\mathbf{C}_{ij}\boldsymbol{\beta}_j$ is $\hat{\boldsymbol{\alpha}}_i^T\mathbf{C}_{ij}\hat{\boldsymbol{\alpha}}_j$, which can be expressed as a quadratic form

$$\hat{\boldsymbol{\alpha}}_i^T\mathbf{C}_{ij}\hat{\boldsymbol{\alpha}}_j = [\hat{\boldsymbol{\alpha}}_i^T \ \hat{\boldsymbol{\alpha}}_j^T]\begin{bmatrix}\mathbf{0} & \frac{1}{2}\mathbf{C}_{ij}\\ \frac{1}{2}\mathbf{C}_{ij}^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\alpha}}_i\\ \hat{\boldsymbol{\alpha}}_j\end{bmatrix} \tag{7}$$

However, this is biased as it can be shown from quadratic form theory that

$$\begin{aligned}\mathrm{E}[\hat{\boldsymbol{\alpha}}_i^T\mathbf{C}_{ij}\hat{\boldsymbol{\alpha}}_j] &\approx \boldsymbol{\beta}_i^T\mathbf{C}_{ij}\boldsymbol{\beta}_j + \frac{\sigma_e^2}{n}\mathrm{tr}\left(\begin{bmatrix}\mathbf{0} & \frac{1}{2}\mathbf{C}_{ij}\\ \frac{1}{2}\mathbf{C}_{ij}^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_i^{-1} & \mathbf{V}_i^{-1}\mathbf{C}_{ij}\mathbf{V}_j^{-1}\\ \mathbf{V}_j^{-1}\mathbf{C}_{ij}^T\mathbf{V}_i^{-1} & \mathbf{V}_j^{-1}\end{bmatrix}\right)\\ &= \boldsymbol{\beta}_i^T\mathbf{C}_{ij}\boldsymbol{\beta}_j + \frac{\sigma_e^2}{n}\mathrm{tr}(\mathbf{C}_{ij}\mathbf{V}_j^{-1}\mathbf{C}_{ij}^T\mathbf{V}_i^{-1}).\end{aligned} \tag{8}$$

Therefore, an approximately unbiased estimator for $\boldsymbol{\beta}_i^T\mathbf{C}_{ij}\boldsymbol{\beta}_j$ can be obtained by subtracting $\frac{\sigma_e^2}{n}\mathrm{tr}(\mathbf{C}_{ij}\mathbf{V}_j^{-1}\mathbf{C}_{ij}^T\mathbf{V}_i^{-1})$ from $\hat{\boldsymbol{\alpha}}_i^T\mathbf{C}_{ij}\hat{\boldsymbol{\alpha}}_j$.

Shrinking $\hat{\boldsymbol{\beta}}^T\hat{\mathbf{V}}^\dagger\hat{\boldsymbol{\beta}}$ towards $\frac{p}{n}$

The quadratic term $\hat{\boldsymbol{\beta}}^T\hat{\mathbf{V}}^\dagger\hat{\boldsymbol{\beta}}$ is usually inflated due to noise generated by eigenvectors of $\hat{\mathbf{V}}^\dagger$ with corresponding small eigenvalues. Thus, a regularization on $\hat{\mathbf{V}}^\dagger$ is necessary to remove the noise. The regularized quadratic term has the general form

$$\hat{\boldsymbol{\beta}}^T(c\hat{\mathbf{V}} + d\mathbf{M})^\dagger\hat{\boldsymbol{\beta}}. \tag{9}$$

Obviously, as $\frac{d}{c}$ goes to infinity, Equation (9) converges to

$$\frac{1}{d}\hat{\boldsymbol{\beta}}^T \mathbf{M}^\dagger \hat{\boldsymbol{\beta}}. \tag{10}$$

Following ridge-regression-style penalty, one sets $\mathbf{M} = \mathbf{I}$, and as $\frac{d}{c} \to \infty$ the quadratic term is shrunk towards $\frac{1}{d}\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}}$. This shrinkage can be problematic as at large $\frac{d}{c}$, it ignores the effect of LD matrix and can result in an upward bias. A more conservative target of shrinkage is $\frac{p}{n}$, the expected value of $\hat{\boldsymbol{\beta}}^T \hat{\mathbf{V}}^\dagger \hat{\boldsymbol{\beta}}$ when there is no heritability.

Ignoring the constant $\frac{1}{d}$ and assuming $\mathbf{M}$ is invertible, we aim to find an $\mathbf{M}$ with the property that

$$\hat{\boldsymbol{\beta}}^T \mathbf{M}^{-1} \hat{\boldsymbol{\beta}} = \frac{p}{n}. \tag{11}$$

Let $\mathbf{M}^{-1} = \mathbf{N}^T\mathbf{N}$, we can then write Equation (11) as

$$\hat{\boldsymbol{\beta}}^T \mathbf{N}^T \mathbf{N} \hat{\boldsymbol{\beta}} = \|\mathbf{N}\hat{\boldsymbol{\beta}}\|_2^2 = \left\| \sum_{i=1}^{p} \hat{\boldsymbol{\beta}}_i \mathbf{N}_i \right\|_2^2 = \frac{p}{n}, \tag{12}$$

where $\mathbf{N}_i$ denotes the i-th column of $\mathbf{N}$. Clearly, when

$$\mathbf{N}_i = \frac{1}{\hat{\boldsymbol{\beta}}_i \sqrt{n}} \mathbf{1}, \tag{13}$$

where $\mathbf{1}$ is the vector of p 1's, Equation (12) is satisfied. Notice that $\hat{\boldsymbol{\beta}}_i \sqrt{n}$ is nothing more than the Z-score (Z_i) of SNP i. Therefore, the matrix

$$\mathbf{M} = \left(\begin{bmatrix} \frac{1}{Z_1} & \cdots & \frac{1}{Z_1} \\ \vdots & \ddots & \vdots \\ \frac{1}{Z_p} & \cdots & \frac{1}{Z_p} \end{bmatrix} \begin{bmatrix} \frac{1}{Z_1} & \cdots & \frac{1}{Z_p} \\ \vdots & \ddots & \vdots \\ \frac{1}{Z_1} & \cdots & \frac{1}{Z_p} \end{bmatrix} \right)^{-1} \tag{14}$$

furnishes a regularization to shrink the quadratic term $\hat{\boldsymbol{\beta}}^T \hat{\mathbf{V}}^\dagger \hat{\boldsymbol{\beta}}$ towards $\frac{p}{n}$.

An example usage of $\mathbf{M}$ is

$$\hat{\boldsymbol{\beta}}^T \left(\frac{1}{1+\lambda} \hat{\mathbf{V}} + \frac{\lambda}{1+\lambda} \mathbf{M} \right)^\dagger \hat{\boldsymbol{\beta}}. \tag{15}$$

As λ goes to infinity, the quadratic term converges to $\frac{p}{n}$.

Adjusting for bias without inverting a matrix

In the second step of the HESS estimator, we need to solve the system of linear equations

$$\hat{h}^2_{g,local,i} = \frac{nf_i - (1 - \sum_{j=1, j\neq i}^{m} \hat{h}^2_{g,local,j})k_i}{n - k_i}, \tag{16}$$

where f_i is the regularized version of $\hat{\boldsymbol{\beta}}_i^T \hat{\mathbf{V}}_i^\dagger \hat{\boldsymbol{\beta}}_i$, and k_i the number of eigenvectors used for locus i. In general, one need to solve a system of linear equation by inverting a $m \times m$ matrix, where m is the number of windows. However, in the special case where $k_1 = \cdots = k_m = k$, inverting a matrix can be avoided.

Using $\hat{h}_g^2$ to denote $\sum_{j=1}^m \hat{h}^2_{g,local,j}$, we note that Equation (16) can be written as

$$\hat{h}^2_{g,local,i} = \frac{nf_i - (1 - \hat{h}_g^2 + \hat{h}^2_{g,local,i})k}{n-k}. \tag{17}$$

Summing over i on both sides gives

$$\begin{aligned}
\hat{h}_g^2 = \sum_{i=1}^m \hat{h}^2_{g,local,i} &= \sum_{i=1}^m \frac{nf_i - (1 - \hat{h}_g^2 + \hat{h}^2_{g,local,i})k}{n-k} \\
&= \frac{n}{n-k}\sum_{i=1}^m f_i - \frac{k}{n-k}\sum_{i=1}^m (1 - \hat{h}_g^2 + \hat{h}^2_{g,local,i}) \\
&= \frac{n}{n-k}\sum_{i=1}^m f_i - \frac{k}{n-k}(m - m\hat{h}_g^2 + \hat{h}_g^2).
\end{aligned} \tag{18}$$

Solving for $\hat{h}_g^2$ gives

$$\hat{h}_g^2 = \frac{n\sum_{i=1}^m f_i - mk}{n-km}. \tag{19}$$

Plugging Equation (19) into Equation (17) and solving for $\hat{h}^2_{g,local,i}$ gives

$$\hat{h}^2_{g,local,i} = \frac{nf_i - k + k\hat{h}_g^2}{n} = f_i - \frac{k}{n}(1 - \hat{h}_g^2). \tag{20}$$

Equation (19) also tells how sensitive the estimate is with respect to the chosen threshold. The derivative of $\hat{h}_g^2$ with respect to the term $\sum_{i=1}^m f_i$ is $\frac{n}{n-mk}$. Therefore, to ensure stability, one should choose m and k such that $n - km$ is large (i.e. $\frac{n}{n-mk}$ is small) so that changes in $\sum_{i=1}^m f_i$ have small effect on the estimates.