1 The Model

This is the model to the best of my knowledge. There are likely errors.

$i \in 0, ..., N$	individuals
$j \in 0, ..., M$	variants
$k \in 0, ..., K$	phenotypes
$D = W \cdot H \cdot C$	image height, width, and number of channels
$h_k^2 = 0.5$	heritability of phenotype k
$p_j \sim \text{Uniform}(0.05, 0.95)$	alternate allele frequency
$X_{ij} \sim \text{Binomial}(2, p)$	genotype
$\beta_{jk} \sim \text{Normal}(0, h^2/M)$	true effect size
$\ell_{ik} \sim X_{ij} \cdot \beta_{jk} + \text{Normal}(0, \sqrt{1 - h^2})$	latent phenotype
$f : \mathbb{R}^K \to \mathbb{R}^D$	an unknown non-linear function
$y_{ik} = f(\ell_{ik})$	image phenotype

The unknown non-linear function models for the complex biological and measurement processes that produce medical images.

2 GWAS Background

Standard GWAS independently tests each variant against the phenotype. Consider, for example, one phenotype y, a column vector with one entry per sample. Let x_j be the j-th column of X_{ij}. Moreover, we will mean center and variance normalize (make the variance 1) the phenotypes and the x_js. Mean centering reduces the degrees of freedom from two to one. Standard linear GWAS fits these M one-degree-of-freedom linear models:

$$y = x_j \beta_j$$

We can solve this with the normal equations. Note that, because x_j is a column vector $x_j^T x_j$ is a real number.

$$\begin{aligned} y &= x_j \beta_j \\ x_j^T y &= x_j^T x_j \beta_j \\ \frac{x_j^T y}{x_j^T x_j} &= \beta_j \end{aligned}$$

Notice that we project the variants onto the phenotypes in N-dimensional space, then we normalize by the square of the magnitude of x_j. However, $x_j^T x_j = 1$ because we mean centered and variance normalized x_j (compare the definition of variance to the definition of vector dot-product). We further simplify the above formula:

$$x_j^T y = \beta_j$$

We can solve for all M betas in one matrix multiplication:

$$X^T y = \beta$$

Finally, if we had multiple phenotypes, we can form a matrix of phenotypes Y and solve for the M by K matrix of betas, B, in one matrix multiply:

$$X^T Y = B$$

3 Test Setup

For testing purposes, we use the Balding-Nichols (BN) model of genotypes with one population. BN generates independent variants. We use ldsc-sim to generate independent phenotypes according to the model. We use the generator from a Deep Convolutional Generative Adversarial Network (DCGAN) as the unknown non-linear function. The generator we used was trained on the CelebA dataset of face images. The latent space of the DCGAN is $\mathbb{R}^{10}$. We choose a random linear transformation to embed lower dimensional latent phenotypes in $\mathbb{R}^{10}$.

4 Trace Heritability

The term "trace heritability" appears throughout the code. This refers to the trace of a heritability matrix for vector phenotypes. I do not fully understand why, but the diagonal of this matrix is just h_k^2 for each phenotype k. Therefore the "trace heritability" is just the sum of h_k^2 over all k.

We include a bias correction term in our *estimate* of the trace heritability because we are squaring β. Alex B asserted our estimates of beta are distributed as described below, but I do not understand why.

$$
\begin{aligned}
\widehat{\beta} &\sim N\left(\beta, \frac{1-\beta^2}{N}\right) \\
\mathbb{E}\widehat{\beta}^2 &= (\mathbb{E}\widehat{\beta})^2 + \mathrm{Var}(\widehat{\beta}) \\
&= \beta^2 + \frac{1-\beta^2}{N} \\
&= \left(1 - \frac{1}{N}\right)\beta^2 + \frac{1}{N}
\end{aligned}
$$

Solve for β^2 to get an unbiased estimator:

$$
\mathbb{E}\widehat{\beta^2} = \frac{N}{N-1}\widehat{\beta}^2 - \frac{1}{N-1}
$$

In the code, I elided the $\frac{N}{N-1}$. I don't think it will have a large effect, but I suppose we should fix that.

In the code, you'll notice the implementation is actually the following because the bias accumulates across K phenotypes and M variants.

$$
\mathbb{E}\widehat{\beta^2} = \widehat{\beta}^2 - \frac{KM}{N-1}
$$

5 Variance Explained

If $K = 1$ then we can easily compare the simulated phenotypes to the latent phenotypes. Plot the simulated and latent phenotype for each individual as a point on a Cartesian plane. This is a scatter-plot. Calculate the Pearson correlation coefficient. The square of the Pearson correlation coefficient is a number in $[0, 1]$ representing the degree of linear correlation between the datasets. If the square of the Pearson correlation coefficient is one, then the datasets are exactly the same. The square of the Pearson correlation coefficient is also known as "R-squared", R^2, or the "coefficient of determination". The "coefficient of determination" is the percent of variance in a dependent variable explained by an independent variable.

If $K > 1$, the scatter-plot interpretation breaks down because we have too many dimensions. Instead, we can take the dual view: a matrix of N K-dimensional observations defines a K-dimensional hyperplane in N dimensional space (if $K < N$). We have two datasets and therefore two hyper-planes in $\mathbb{R}^N$. These hyper-planes meet at K angles. The square of the cosine of the angles gives us a number in $[0, 1]$. If the planes are equal, then every angle is zero and every square-cosine is 1. If we sum the square-cosines for each angle and divide by K then we have a measure in $[0, 1]$ of the similarity between the two hyperplanes.

For mean centered, variance normalized phenotypes and genotypes, we can connect the Pearson correlation coefficient and the coefficient of determination to the angle between two N-dimensional vectors. The Pearson correlation coefficient is defined for a pair of scalar-valued datasets x and y as:

$$r_{xy} = \frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\text{var}(x)\text{var}(y)}}$$

If x and y are mean centered and variance normalized column vectors, then this formula simplifies:

$$r_{xy} = \sum_i x_i y_i = x^T y = \widehat{\beta}$$

The coefficient of determination is the percent of variance in y explained by some prediction z:

$$R^2_{xy} = 1 - \frac{\sum_i (y_i - z_i)^2}{\text{var}(y)}$$

Again, for mean centered and variance normalized vectors, this formula simplifies:

$$\begin{aligned}
R^2_{xy} &= 1 - \frac{\sum_i (y_i - z_i)^2}{\text{var}(y)} \\
&= 1 - \sum_i (y_i - z_i)^2 \\
&= 1 - (y - z)^T (y - z) \\
&= 1 - (y^T (y - z) - z^T (y - z)) \\
&= 1 - (y^T y - y^T z - z^T y + z^T z) \\
&= 1 - y^T y + y^T z + z^T y - z^T z) \\
&= 1 - 1 + y^T z + z^T y - z^T z && \text{y is mean centered and has variance one} \\
&= y \cdot z + z \cdot y - z \cdot z && \text{convert matrix ops to vector dot-products} \\
&= 2y \cdot z - z \cdot z && \text{commutativity of dot-product}
\end{aligned}$$

When using a linear model, the prediction, z, in a coefficient of determination is just x $\widehat{\beta}$. We substitute that and simplify:

$$
\begin{aligned}
R^2_{xy} &= 2y \cdot z - z \cdot z \\
&= 2y \cdot \widehat{\beta} x - \widehat{\beta}^2 x \cdot x \\
&= 2y \cdot \widehat{\beta} x - \widehat{\beta}^2 && \text{x is mean centered and has variance one} \\
&= 2\widehat{\beta} y \cdot x - \widehat{\beta}^2 && \beta \text{ is a scalar} \\
&= 2\widehat{\beta}^2 - \widehat{\beta}^2 && y \cdot x = y^T x = \widehat{\beta} \\
&= \widehat{\beta}^2 \\
&= r^2_{xy}
\end{aligned}
$$

I imagine there is a less algebraic and more intuitive explanation for this, but I do not know it. The final quantity we want to relate is the angle between two vectors in N-dimensional space. Recall that our datasets contain scalar observations so we can view each one as an N-dimensional vector. Two N-dimensional vectors have an angle, θ, which is related to the dot-product of these vectors:

$$x \cdot y = ||x||||y||\cos(\theta)$$

Recall that the magnitude of x is the square root of $x \cdot x$. Again, due to mean centering and variance normalizing, our magnitudes are all 1.

$$
\begin{aligned}
x \cdot y &= \cos(\theta) \\
x^T y &= \cos(\theta) \\
\widehat{\beta} &= \cos(\theta)
\end{aligned}
$$

Collecting our identities for the linear model $y = x\beta$:

$$
\begin{aligned}
R^2_{xy} &= (r_{xy})^2 \\
r_{xy} &= x^T y \\
x^T y &= \widehat{\beta}_{xy} \\
\widehat{\beta}_{xy} &= \cos(\theta_{xy})
\end{aligned}
$$

Recall that R^2 defines the percent of variance in y explained by x. More importantly, this generalizes to non-scalar datasets like our K-dimensional latent phenotypes. In the higher-dimensional setting, a factor of K appears and we have K angles between the K-dimensional hyper-planes rather than one angle between the vectors (1-dimensional hyperplanes).

$$
\begin{aligned}
X &: \mathbb{R}^{N \times M} \\
Y &: \mathbb{R}^{N \times K} \\
B &: \mathbb{R}^{K}
\end{aligned}
$$

$$
\begin{aligned}
R_{xy}^2 &= \frac{1}{K}(r_{xy})^2 \\
r_{xy} &= ||X^T Y||_F \\
X^T Y &= \widehat{B}_{xy} \\
||\widehat{B}_{xy}||_F &= \sum_{k=0}^{K} \cos(\theta_k)
\end{aligned}
$$

I think there is an error in the above equations. I feel like the factor of K should be a factor of K^2. We should talk to Alex B about this.