Same eigen decomposition: proof

Proof for the same pearson correlation coefficient

Let us note x and y two SNPs and $\overline{x}$ the mean of x.

Then the pearson correlation coefficient between x and y is:

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \quad .$$

If we note x_K and y_K the SNPs x and y replicated K times, then, $\overline{x_K} = \frac{K \sum_{i=1}^{n} x_i}{K\ n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$ and r_K the pearson correlation coefficient between x_K and y_K is

$$r_K = \frac{K \sum_{i=1}^{n} (x_i - \overline{x_K})(y_i - \overline{y_K})}{\sqrt{K \sum_{i=1}^{n} (x_i - \overline{x_K})^2} \sqrt{K \sum_{i=1}^{n} (y_i - \overline{y_K})^2}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} = r \ .$$

Proof for the same eigen analysis

First, as $\overline{x_K} = \overline{x}$, then the allele frequency of x_K is the same as for x so that the scaling $\frac{x-2p}{\sqrt{2p(1-p)}}$ is also the same (because $p = \overline{x}/2$).

For the exact singular value decomposition of $G = U \Delta V^T$ (where G is the scaled genotype matrix), we can first compute $\Sigma = G^T G = V \Delta^2 V^T$, then remark that $\Sigma V = V \Delta^2$ so that V is the matrix of the eigen vectors of Σ and Δ^2 is the matrix of the eigen values of Σ. Finally to get U, we can compute $G V \Delta^{-1} = U \Delta V^T V \Delta^{-1} = U \Delta \Delta^{-1} = U$.

For replicated individuals, we want the decomposition of $G_K = U_K \Delta_K V_K^T$. Then, $\Sigma_K = G_K^T G_K = K G^T G = K \Sigma$ so that $V_K = V$ (same PC loadings) and $\Delta_K^2 = K \Delta^2$ resulting in $\Delta_K = \sqrt{K} \Delta$ (same eigen values, up to a constant). Finally $U_K = G_K V_K \Delta_K^{-1} = G_K V \left(\sqrt{K} \Delta\right)^{-1} = \frac{1}{\sqrt{K}} G_K V \Delta^{-1}$ (PCs scores are the same, up to a constant).