

I generate PCA for all samples in the reference databases after LD filtration and calculated projection matrix by  $M'V'\Sigma^{-2}$ , then plot target sample on this PCA.

I assume that any sample is either from a single population or a linear mixture of different populations and can be represented by the following linearly additive model (P represents each population, x represents the proportions and  $R_i$  represents the discrepancy of the observed value to the expected value):

$$\begin{aligned} S_i &= \sum_j P_j x_{ij} + R_i \\ &= \left( P_1 \dots P_j \right) \begin{pmatrix} x_{i1} \\ \dots \\ x_{ij} \end{pmatrix} + R_i \quad \textcircled{1} \\ S_i &= PX_i + R_i \end{aligned}$$

and the variances should also follow linearly additive model.

$$V_i = (\sigma_1 \dots \sigma_j) X_i = \sigma X_i \quad \textcircled{2}$$

Thus, combining the ① and ②, the discrepancy relative to variance can be represented by

$$\begin{aligned} S_i &= PX_i + R_i \\ &= PX_i + \text{diag}(V_i) R_i' \quad \textcircled{3} \\ \text{diag}^{-1}(V_i) S_i &= \text{diag}^{-1}(V_i) PX_i + R_i' \end{aligned}$$

While treating  $\text{diag}^{-1}(V_i)$  as a constant matrix, and assuming that regardless with or without relative to the variance,  $S_i$ 's nearest point in space M should be close enough to each other.

Assuming the discrepancies correspond to the shortest distances to vector spaces formed by linear additive of super-populations, the discrepancy of each sample to any combination of populations can be determined as the following:

$$\begin{aligned} \text{Any sample } \text{diag}^{-1}(V_i) S_i \text{ 's shortest distance to space } \text{diag}^{-1}(V_i) P \text{ is the} \\ \text{Re } jction_{S_i \rightarrow P} = \text{diag}^{-1}(V_i) S_i - \text{Pr } ojection_{S_i \rightarrow P} = \text{diag}^{-1}(\sigma X_i) (I - P(P^T P)^{-1} P^T) S_i \end{aligned}$$

$$\text{And any sample's ethnicity composition is estimated as } X_i = (P^T P)^{-1} P^T S_i.$$

All samples' ethnicity compositions were estimated and minor components (proportion < 0.01) were discarded.