**Coancestry and individual allele frequencies**

Let \( \pi_{ij} \) denote the individual allele frequency (IAF) of individual \( j \) at SNP \( i \). The IAF moments are

\[
E(\pi_{ij}) = p_i, \\
\text{Cov}(\pi_{ij}, \pi_{jk}) = p_i(1 - p_i)\theta_{jk}, \\
x_{ij}/\pi_{ij} \sim \text{Binomial}(2, \theta_{ij}),
\]

where \( \theta_{jk} \in [0, 1] \) are individual coancestry coefficients. This models locally outbred and locally unrelated individuals, and generalizes the model of [5]. Under these assumptions, kinship and coancestry are the same:

\[
\theta_{jk} = \frac{\sum w_{jk} - 1}{\theta_{jj}} \quad \text{if} \quad j \neq k, \\
\theta_{jj} = \frac{1}{\theta_{jj}} \quad \text{if} \quad j = k.
\]

We also have an \( F_{ST} \) analogous to a previous definition [3]

\[
F_{ST} = \sum w_{jk}\bar{\theta}_{jk}.
\]

**Inconsistency in MM estimators**

The "naïve" MM coancestry estimator is, and converges to,

\[
\hat{\theta}_{jk} = \sum w_{jk}(\pi_{ij} - \bar{\theta}_{ij}) \sim \frac{\sum w_{jk} - 1}{\theta_{jj}} \theta_{jk} - \theta_{ij} + \theta
\]

as the number of SNPs \( m \to \infty \), where \( \bar{\theta}_{ij} = \sum w_{jk} \theta_{jk} = \sum w_{jk} \theta_{jk} \) and \( \theta = \sum \sum w_{jk} \theta_{jk} \). So these estimates suffer from column- and row-specific distortions. The genotype version \( \pi_{ij} \to x_{ij}/2 \) is a popular kinship estimator with similar biases.

The "naïve" \( F_{ST} \) estimator by \( \hat{\theta}_{jk} \) is, and converges to,

\[
\hat{F}_{ST} = \sum w_{jk}\bar{\theta}_{jk} = \sum w_{jk} (\pi_{ij} - \bar{\theta}_{ij})^2 \sim \frac{F_{ST} - 1}{\theta_{jj}} - 1
\]

analogous to a previous result for populations [5]. Since 0 < \( \theta \leq F_{ST} \), \( \hat{F}_{ST} \) may be arbitrarily close to zero, even for large true \( F_{ST} \). In practice \( \theta \) is unknown.

**Simulation results**

We constructed an admixture simulation that induces extreme biases in existing \( F_{ST} \) estimates (fig. 1). While the WC and Hudson \( F_{ST} \) estimators are unbiased under their respective models, they are indeed severely biased in our admixture model (fig. 2).

Our simulation also illustrates the downward bias and gross distortions of estimated coancestries (and kinships) using the MM approach (fig. 3).

---

**References**

