

Multi-allelic generalization of the numerically stable computation of the Exact AF calculation

Let N_{alt} = number of alternate alleles; let $K = \sum_{i=1}^{N_{alt}} k_i$; and let $\alpha(k_{i-n}) = k_1, \dots, k_{i-n}, \dots, k_{N_{alt}}$

Then in the case of diploid samples:

$$y_{j,k_1,\dots,k_{N_{alt}}} = \frac{1}{2j(2j-1)} \cdot \left[\begin{array}{l} (2j-K)(2j-K-1) \cdot y_{j-1,k_1,\dots,k_{N_{alt}}} \cdot P(D_j | \langle aa \rangle) \\ + \sum_{i=1}^{N_{alt}} 2k_i(2j-K) \cdot y_{j-1,\alpha(k_i-1)} \cdot P(D_j | \langle aA_i \rangle) \\ + \sum_{i=1}^{N_{alt}} k_i(k_i-1) \cdot y_{j-1,\alpha(k_i-2)} \cdot P(D_j | \langle A_iA_i \rangle) \\ + \sum_{i=1, l=1, i < l}^{N_{alt}} 2k_i k_l \cdot y_{j-1,\alpha(k_i-1, k_l-1)} \cdot P(D_j | \langle A_iA_l \rangle) \end{array} \right]$$

(Hom Ref)

(Het Ref)

(Hom Var)

(Het Var)