

## Allele sharing for mixtures

Consider marker  $i$  with mixture  $M_i = (A_{i1}, \dots, A_{iI})$  and corresponding allele frequencies  $p_{i1}, \dots, p_{iI}$ . The number of alleles the defendant shares with the mixture for this marker is denoted  $Z_i$ . Let  $S_i = p_{i1} + \dots + p_{iI}$  be the sum of the allele frequencies at marker  $i$ . Then a direct argument gives (calculations assume HWE and  $H_d$ )

$$P(Z_i = 0) = (1 - s_i)^2$$

$$P(Z_i = 1) = 2s_i(1 - s_i)$$

$$P(Z_i = 2) = s_i^2$$

Let  $Z = Z_1 + \dots + Z_I$  be the total number of alleles shared and  $\mathbf{w} = (w_1, \dots, w_I)$  where  $w_i \in \{0, 1, 2\}$  one of these values. Then for a  $k = 0, \dots, I, \dots, 2I$ ,

$$P(Z = k) = \sum_{\text{all permutations in } \mathbf{w}: \sum w_i = k} \prod_{i=1}^I P(Z_i = w_i)$$

Here “all permutations” means all possible ordered combinations of the elements in the vector  $\mathbf{w}$ . Note here that RMNE simplifies to  $P(Z = 2I) = \prod_{i=1}^I P(Z_i = 2)$ .

## Implementation:

Consider that we want to calculate  $P(Z = 16)$ , where  $I = 10$

- 1) First, calculate  $P(Z_i = w_i)$  for each locus  $i$  and  $w_i \in \{0, 1, 2\}$
- 2) Let  $x = \{w_1, w_2, \dots, w_9, w_{10}\}$  a vector with 0, 1 or 2 elements
- 3) Find all the possible sequences such that  $\text{sum}(x) = 16$  (by recursion) and  $\text{length}(x) \leq 10$  by recursion
  - a. 247 sequences satisfy this.
- 4) Find the unique sequences where the loci-order does not matter (combinations), and zero is removed
  - a. 3 unique sequences satisfies this:
    - i. 1 1 1 1 2 2 2 2 2 2
    - ii. 1 1 2 2 2 2 2 2 2 2
    - iii. 2 2 2 2 2 2 2 2 2 2
- 5) For each of the 3 sequences:
  - a. Permute the order of the digits in each of the 3 unique sequences to obtain  $L$  permuted sequences.
    - i.  $L=210$  for i.
    - ii.  $L=360$  for ii.

- iii.  $L=45$  for iii.
  - b. From this we obtain a  $(L \times I)$  matrix  $A$  with elements 0, 1 or 2.
  - c. Construct another  $(L \times I)$  matrix  $B$  with elements
    - i.  $B_{l,i} = P(Z_i = A_{l,i})$
  - d. From this, calculate  $S = \sum_l \prod_i B_{l,i}$ 
    - i. Note:  $S = \sum_l \exp(\sum_i \log B_{l,i})$
- 6) Sum  $S$  together for each of the 3 unique sequences to calculate  $P(Z = 16)$