A. Constraints on individual LD values

The normalized individual LD values are \( D_i' = D_i / D_{\text{max}} \), where
\[
D_{\text{max}} = \min[p_{AB}(1-p_B), (1-p_A)p_B] \text{ if } D_i > 0, \quad \text{and}
\]
\[
D_{\text{max}} = \min[p_{AB}p_B(1-p_B)(1-p_B)] \text{ when } D_i < 0. \quad \text{The range of } D_i' \text{ is } (-1, 1).
\]

B. Other measures of conditional LD

The conditional association between the alleles at marker locus B and disease locus A was defined by Nei and Li (1980) as
\[
d^* = (f_{A2B1} - f_{B2A1}) / (f_{B2A1} + f_{A2B1})
\]
for two bi-allelic loci. It was developed to account for study designs for rare diseases with known modes of inheritance and full, or close to full, penetrance (e.g., sickle cell anemia and Duchenne muscular dystrophy) where individuals are not randomly sampled from a single population. This measure is equivalent to Somer’s D statistic, \( D(C|R) \), conditioning on the rows of the contingency table relating two categorical variables, as shown below.
\[
d_i = \frac{f_{A2B1}/p_{A2} - f_{B2A1}/p_{B1}}{p_{A2} - (p_{A1}p_{B1} - D)/p_{A1}}
\]
\[
= p_{B1} + (D/p_{A1}) - p_{A1} + (D/p_{A1}) = D / (p_{A1}p_{B2})
\]
Somer’s D statistic is defined as twice the difference between the number of concordant and discordant entries in the contingency table, divided by \( n^2 \cdot \frac{2}{n^2} \), where \( n \) is the total and \( n_i \) is the \( i \)th row total.
\[
D(C|R) = 2(n_{ij}n_{k\ell} - n_{ij}n_{k\ell}) / (\sum_i n_i^2)^2 = 2(f_{A2B1}f_{B2A1} - f_{A2B1}f_{B2A1}) / (1 - \sum_i p_i^2)
\]
\[
= 2D(1-F_A) = 2D / (2p_{A2}p_{B2}) = D / (p_{A1}p_{B2})
\]

C. Other measures of LD based on diversity statistics

Other measures of association and LD that are based on allelic diversity statistics (homozygosity and heterozygosity) have been defined. However, these measures are all symmetric. Ohta (1980) suggested a measure, \( F' \), that divides the difference between the two-locus haplotypic homozygosity \( (F_{AB}) \) and the product of the two single locus homozygosity values by the product of the single locus heterozygosity values:
\[
F' = (f_{AB} - F_A F_B) / [(1 - F_A)(1 - F_B)] \quad \text{(Ohta 1980)}.
\]
The \( D^* \) measure (Maruyama 1982; Hedrick and Thomson 1986) standardizes the disequilibrium between all alleles at two loci by the product of the single locus heterozygosity values:
\[
D^* = \Sigma_i \Sigma_j D_{ij} / [(1 - F_A)(1 - F_B)] \quad \text{Note that in the bi-allelic case } D^* \text{ is equivalent to the square of the correlation measure } (r^2) \quad \text{(Hedrick 1987)}.
\]

D. Proofs of Special Cases (c) – (f)

D.1. Multi-Allelic Case (c): \( k_A = k_B = k, f(AiBi) > 0, i = 1, 2, ..., k, \) and \( f(AiBj) = 0 \) for all \( i \neq j \)

**Summary:** \( W_{Ai} = W_{AiB} = W_{BiA} = 1 \). There is complete symmetry and 100% correlation of alleles at the two loci.

The A and B locus allele frequencies are equal and the notation is simplified as follows: \( p_{Al} = p_{Bl} = p_i, i = 1, 2, ..., k, \)
\[
f(AiBj) = p_i p_j + D_{ij} \quad \text{i.e., } D_{ij} = p_i(1-p_i), f(AiBj) = 0 = p_i p_j + D_{ij} \quad \text{i.e., } D_{ij} = - p_i p_j, i \neq j.
\]
\[
W_{Ai}^2 = (\Sigma_i \Sigma_j [f_{ij} / p_{ij}] - F_A) / (1 - F_A) = ([\Sigma_i f_{ij}^2 / p_{ij}] - F_A) / (1 - F_A)
\]
Similarly, \( W_{AB}^2 \) = 1.

\[
W_i^2 = [\sum_i D_i^2 / (p_{AI} p_{Bj})] / (k-1) = [\sum_i D_i^2 / (p_i^2)] / (k-1) = [\sum_i (1 - p_i^2) + \sum_k (1 - p_i k)] / (k-1) = [\sum_i (1 - p_i^2) + \sum_k (1 - p_i k)] / (k-1)
\]

D.2. Multi-Allelic Case (d): \( k_A = k_B = k, f(AIB) > 0, i = 1, 2, ..., k, \) and \( f(A1B2) \neq 0 \)

In this case \( W_{A/B} < 1, W_{B/A} < 1, \) and \( W_A < 1 \) (Proof given below for \( k = 3, \) with the same result holding for any value of \( k \)).

Summary: \( W_{A/B}^2 = (1 - 2 \delta [p_{A2} / p_{A2}] - F_A) / (1 - F_A), W_{B/A}^2 = (1 - 2 \delta [p_{A1} / p_{A3}] - F_B) / (1 - F_B), W_A^2 = 1 - \delta (p_{A1} + p_{A2}) / (2 p_{A1} p_{A2}) \)

Allele frequencies are \( (A \) locus): \( p_{A1}, p_{A2}, \) and \( p_{A3} ; (B \) locus): \( p_{B1}, p_{B2}, \) and \( p_{B3} \), with \( \delta > 0 \) and \( \delta < p_{A1} \).

<table>
<thead>
<tr>
<th>Haplotypes</th>
<th>Frequency</th>
<th>Frequency*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1</td>
<td>( p_{A1} - \delta )</td>
<td>( p_{A1} + (p_{A2} - \delta) + D_{12} )</td>
</tr>
<tr>
<td>A1B2</td>
<td>( \delta )</td>
<td>( p_{A1} + (p_{A2} + \delta) + D_{12} )</td>
</tr>
<tr>
<td>A1B3</td>
<td>0</td>
<td>( p_{A2} + D_{13} )</td>
</tr>
<tr>
<td>A2B1</td>
<td>0</td>
<td>( p_{A2} + D_{21} )</td>
</tr>
<tr>
<td>A2B2</td>
<td>( p_{A2} )</td>
<td>( p_{A2} + D_{22} )</td>
</tr>
<tr>
<td>A2B3</td>
<td>0</td>
<td>( p_{A3} + D_{23} )</td>
</tr>
<tr>
<td>A3B1</td>
<td>0</td>
<td>( p_{A3} + (p_{A2} - \delta) + D_{31} )</td>
</tr>
<tr>
<td>A3B2</td>
<td>0</td>
<td>( p_{A3} + (p_{A2} + \delta) + D_{32} )</td>
</tr>
<tr>
<td>A3B3</td>
<td>( p_{A3} )</td>
<td>( p_{A3} + D_{33} )</td>
</tr>
</tbody>
</table>

* These two columns are different, but equivalent, formats for the same haplotype frequency.
\[ + p_{A2} (p_{A2} + \delta) + p_{A2} p_{A3} + p_{A3} (p_{A1} - \delta) + p_{A3} (p_{A2} + \delta) + (1 - p_{A3})^2) / 2 \]

\[ = [(1 - p_{A1})^2 - \delta (1 - p_{A2})^2 / p_{A1} + p_{A1} p_{A2} - 2 \delta + \delta^2 / p_{A1} (p_{A2} + \delta)] + p_{A1} p_{A3} + p_{A2} p_{A1} - \delta p_{A2} + [p_{A2} / (p_{A2} + \delta)] \]

\[ + [1 - (p_{A2} + \delta) / (p_{A2} + \delta)] - 2 p_{A2} + \delta p_{A2} + p_{A2} p_{A3} + p_{A3} - 6 p_{A3} + p_{A2} + 6 p_{A3} + (1 - p_{A3})^2) / 2 \]

\[ = [(1 - p_{A1})^2 + 2 p_{A1} p_{A2} + 2 p_{A2} p_{A3} + 2 p_{A2} p_{A3} + (1 - p_{A2})^2 + (1 - p_{A2})^2 - \delta (1 - p_{A1})^2 / [p_{A1} + \delta p_{A1} - 2 \delta + \delta^2 / [p_{A1} (p_{A2} + \delta)] - [\delta / (p_{A2} + \delta)] / 2 \]

\[ = (3 - 2 (p_{A1} + p_{A2} + p_{A3}) + (p_{A1} + p_{A2} + p_{A3})^2 - [\delta / p_{A1}] + 2 \delta p_{A1} + 6 p_{A1} - 2 \delta + \delta^2 / [p_{A1} (p_{A2} + \delta)] - [\delta / (p_{A2} + \delta)] / 2 \]

\[ = 1 - [\delta (p_{A1} + p_{A2}) / [2 p_{A1} p_{A2}]] < 1 \text{ always since} \delta > 0. \]

**D.3. Multi-Allelic Case (e):} k_A < k_B, \text{ with each} B_j \text{ allele occurring with only one} A_i \text{ allele}\

In this case \( W_n = W_{A/B} = 1 \), and \( W_{B/A} < 1 \) (Proof given below for \( k_A = 3 \) and \( k_B = 4 \), and the equivalent result holds if \( k_A > k_B \)).

**Summary:** \( W_{A/B}^2 = 1, \) \( W_{B/A}^2 = \left[1 - [2 p_{B3} p_{B4} / ([p_{B3} + p_{B4}])] - F_3 \right] / (1 - F_3), W_r^2 = 1 \)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency*</th>
<th>Frequency*</th>
<th>LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1</td>
<td>( p_{B1} )</td>
<td>( p_{B1}^2 + D_{11} )</td>
<td>( D_{11} = p_{B1} (1 - p_{B1}) )</td>
</tr>
<tr>
<td>A1B2</td>
<td>0</td>
<td>( p_{B1} p_{B2} + D_{12} )</td>
<td>( D_{12} = - p_{B1} p_{B2} )</td>
</tr>
<tr>
<td>A1B3</td>
<td>0</td>
<td>( p_{B1} p_{B3} + D_{13} )</td>
<td>( D_{13} = - p_{B1} p_{B3} )</td>
</tr>
<tr>
<td>A1B4</td>
<td>0</td>
<td>( p_{B1} p_{B4} + D_{14} )</td>
<td>( D_{14} = - p_{B1} p_{B4} )</td>
</tr>
<tr>
<td>A2B1</td>
<td>0</td>
<td>( p_{B2} p_{B1} + D_{21} )</td>
<td>( D_{21} = - p_{B2} p_{B1} )</td>
</tr>
<tr>
<td>A2B2</td>
<td>( p_{B2} )</td>
<td>( p_{B2}^2 + D_{22} )</td>
<td>( D_{22} = p_{B2} (1 - p_{B2}) )</td>
</tr>
<tr>
<td>A2B3</td>
<td>0</td>
<td>( p_{B2} p_{B3} + D_{23} )</td>
<td>( D_{23} = - p_{B2} p_{B3} )</td>
</tr>
<tr>
<td>A2B4</td>
<td>0</td>
<td>( p_{B2} p_{B4} + D_{24} )</td>
<td>( D_{24} = - p_{B2} p_{B4} )</td>
</tr>
<tr>
<td>A3B1</td>
<td>0</td>
<td>( p_{B1} (p_{B3} + p_{B4}) + D_{31} )</td>
<td>( D_{31} = - p_{B1} (p_{B3} + p_{B4}) )</td>
</tr>
<tr>
<td>A3B2</td>
<td>0</td>
<td>( p_{B2} (p_{B3} + p_{B4}) + D_{32} )</td>
<td>( D_{32} = - p_{B2} (p_{B3} + p_{B4}) )</td>
</tr>
<tr>
<td>A3B3</td>
<td>( p_{B3} )</td>
<td>( p_{B3} (p_{B3} + p_{B4}) + D_{33} )</td>
<td>( D_{33} = p_{B3} (1 - p_{B3} - p_{B4}) )</td>
</tr>
<tr>
<td>A3B4</td>
<td>( p_{B4} )</td>
<td>( p_{B4} (p_{B3} + p_{B4}) + D_{34} )</td>
<td>( D_{34} = p_{B4} (1 - p_{B3} - p_{B4}) )</td>
</tr>
</tbody>
</table>

Allele frequencies at the A locus are: \( p_{A1}, p_{A2}, \) and \( p_{A3} = (p_{B1} + p_{B4}), \) and at the B locus they are: \( p_{B1} (= p_{A1}), p_{B2} (= p_{A2}), p_{B3} \) and \( p_{B4} = p_{B4} \).

\[ W_A/n^2 = \left[\sum_i [f_i^2 / p_{B1}] - F_3 \right] / (1 - F_3) = \left[ p_{B1} + p_{B2} + p_{B3} + p_{B4} - F_3 \right] / (1 - F_3) = 1 \]

\[ W_{B/A}^2 = \left[\sum_i [f_i^2 / p_{B1}] - F_3 \right] / (1 - F_3) \]

\[ = [p_{B1} + p_{B2} + [p_{B3}^2 / (p_{B3} + p_{B4})] + [p_{B4}^2 / (p_{B3} + p_{B4})] - F_3] / (1 - F_3) \]

\[ = [p_{B1} + p_{B2} + [p_{B3}^2 + p_{B4}^2] / (p_{B3} + p_{B4})] - F_3 / (1 - F_3) \]

\[ = [p_{B1} + p_{B2} + [(p_{B3} + p_{B4})^2 / (p_{B3} + p_{B4})] - 2 p_{B3} p_{B4} / (p_{B3} + p_{B4})] / (1 - F_3) \]

\[ = [1 - [2 p_{B3} p_{B4} / (p_{B3} + p_{B4})] - F_3] / (1 - F_3) < 1 \text{ always} \]

\[ W_r^2 = \left[\sum_i D_i^2 / (p_{A1} p_{B1}) \right] / (k_A - 1) \]

\[ = \left(1 - p_{B1}^2\right)^2 + p_{B1} p_{B2} + p_{B1} p_{B3} + p_{B1} p_{B4} + p_{B2} p_{B1} + [1 - p_{B2}^2] p_{B3} + p_{B2} p_{B4} + p_{B3} p_{B1} + [p_{B3} + p_{B4}] + p_{B2} [p_{B3} + p_{B4}] - 2 p_{B3} + p_{B4}] / 2 \]

\[ = (3 - 2 (p_{B1} + p_{B2} + p_{B3}) + (p_{B1} + p_{B2} + p_{A3})^2) / 2 = 1 \]

**D.4. One Locus Bi-Allelic and the Other Multi-Allelic, Case (f):} k_A = 2, k_B > 2.\)
In this case $W_n = W_{A/B} \neq W_{B/A}$ (Proof given below for $k_A = 2$ and $k_B = 3$, with the obvious extension for any value of $k_A$)

Summary: $W_{A/B}^2 = [\{(D_1^2 / (p_{A1})) + (D_11 + D_12)^2 / (p_{B1})\} + (D_11 + D_12)^2 / (p_{B2})]/(p_{A1} p_{A2})$, $W_{B/A}^2 = [(D_{11}^2 + D_{12}^2 + (D_{11} + D_{12})^2] / (p_{A1} p_{A2})(1 - F_b)]$,

$W_n^2 = W_{A/B}^2$

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
<th>LD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1</td>
<td>$p_{A1} p_{A2} + D_{11}$</td>
<td>$D_{11}$</td>
</tr>
<tr>
<td>A1B2</td>
<td>$p_{A1} p_{B2} + D_{12}$</td>
<td>$D_{12}$</td>
</tr>
<tr>
<td>A1B3</td>
<td>$p_{A1} p_{A3} + D_{13}$</td>
<td>$-D_{11} - D_{12}$</td>
</tr>
<tr>
<td>A2B1</td>
<td>$p_{A2} p_{A1} + D_{21}$</td>
<td>$-D_{11}$</td>
</tr>
<tr>
<td>A2B2</td>
<td>$p_{A2} p_{B2} + D_{22}$</td>
<td>$-D_{12}$</td>
</tr>
<tr>
<td>A2B3</td>
<td>$p_{A2} p_{B3} + D_{23}$</td>
<td>$D_{11} + D_{12}$</td>
</tr>
</tbody>
</table>

* See Table 1 footnote for constraints on LD parameters, such that two LD variables, $D_{11}$ and $D_{12}$, along with three allele frequencies ($p_{A1}$, $p_{B2}$, and $p_{B3}$) are sufficient to describe the five independent haplotype frequencies in this case.

$W_{A/B}^2 = [\sum_i D_{ij}^2 / (p_{B0})] / (1 - F_B)$

$= [(D_{11}^2 / (p_{A1})) + (D_{12}^2 / (p_{A2})) + (D_{11} + D_{12})^2 / (p_{B1})] + (D_{12}^2 / (p_{B2})) + (D_{11} + D_{12})^2 / (p_{B3})] / (2 p_{A1} p_{A2})$

$W_{B/A}^2 = [\sum_i D_{ij}^2 / (p_{A0})] / (1 - F_B)$

$= [(D_{11}^2 / (p_{A1})) + (D_{12}^2 / (p_{A2})) + (D_{11} + D_{12})^2 / (p_{B1})] + (D_{12}^2 / (p_{B2})) + (D_{11} + D_{12})^2 / (p_{B3})] / (1 - F_B)$

$W_n^2 = [\sum_i D_{ij}^2 / (p_{A0} p_{B0})] / (2 - 1)$

$= [(D_{11}^2 / (p_{A1})) + (D_{12}^2 / (p_{A2})) + (D_{11} + D_{12})^2 / (p_{A1} p_{B2})] + (D_{11} + D_{12})^2 / (p_{B2})] + (D_{12}^2 / (p_{A2} p_{B1})) + (D_{11} + D_{12})^2 / (p_{A2} p_{B3})]$

$= [(D_{11}^2 / (p_{A1})) + (D_{12}^2 / (p_{A2})) + (D_{11} + D_{12})^2 / (p_{B0})] / (p_{A1} p_{A2}) = W_{A/B}^2$