Measures of Linkage Disequilibrium
Part I: A Development of the LD Measure D'

statgen.org

March 2007

We look at the relationship between 2 distinct **phase-known** loci.

		Allele at 1st locus		
		A	a	
Allele at 2nd locus	B	p_{AB}	p_{aB}	p_B
	b	p_{Ab}	p_{ab}	p_b
		p_A	p_a	1

Table 1: Haplotype Probabilities

We define $D \equiv p_{AB} - p_A p_B$. Note then that $D = p_{AB} p_{ab} - p_{Ab} p_{aB}$:

$$\begin{aligned} D &= p_{AB} - p_A p_B \\ &= p_{AB} - (p_{AB} + p_{Ab})(p_{AB} + p_{aB}) \\ &= p_{AB} - (p_{AB}^2 + p_{AB} p_{aB} + p_{Ab} p_{AB} + p_{Ab} p_{aB}) \\ &= p_{AB}(1 - p_{AB} - p_{aB} - p_{Ab}) - p_{Ab} p_{aB} \\ &= p_{AB} p_{ab} - p_{Ab} p_{aB}. \end{aligned}$$

From the definition of D, it follows that $p_{AB} = p_A p_B + D$. In fact, we can write each entry in the table above using only D and the entry's expected value under the null hypothesis of no association. As an example, we develop the expression for p_{aB}:

$$\begin{aligned} p_B &= p_{AB} + p_{aB} \\ &= p_A p_B + D + p_{aB} \end{aligned}$$

(rewriting by isolating p_{aB} on the left hand side)

$$\begin{aligned} p_{aB} &= p_B(1 - p_A) - D \\ p_{aB} &= p_B p_a - D. \end{aligned}$$

Thus, Table 1 becomes:

		Allele at 1st locus		
		A	a	
Allele at 2nd locus	B	$p_A p_B + D$	$p_a p_B - D$	p_B
	b	$p_A p_b - D$	$p_a p_b + D$	p_b
		p_A	p_a	1

Table 2: Haplotype Probabilities as Functions of D and the Expected Genotype Probabilities Under H_0

The fact that no entry in the table can be negative imposes limits on the possible value of D. For example, if D is positive, we must have that $p_A p_b - D \geq 0$ and $p_a p_B - D \geq 0$ so that $D \leq \min(p_A p_b, p_a p_B)$. Similarly, if D is negative, we must have that $-D \leq \min(p_A p_B, p_a p_b)$.

We may wish to have a measure that is always on the interval $(0, 1)$ and that approaches 1 as the association between the loci increases. If so, we can begin by working with $-D$ if $D < 0$ and D if $D \geq 0$. We can then introduce the variable D_{max} such that:

$$D_{max} = \begin{cases} \min(p_A p_b, \quad p_a p_B), & D \geq 0 \\ \min(p_A p_B, \quad p_a p_b), & D < 0. \end{cases}$$

Finally, with the above in mind, we can define $D' = \frac{|D|}{D_{max}}$, which is always on $(0, 1)$ and that approaches 1 as the association between the loci increases.

References

[1] Thomas, Duncan C., Statistical Methods in Genetic Epidemiology. Oxford University Press, Oxford, UK, 2004.