Measures of Linkage Disequilibrium
Part II: A Development of the LD Measure R

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 $D \equiv p_{AB} - p_A p_B$ and 2 random variables (RVs) X and Y such that

$$X = \begin{cases} 1, & \text{Allele 1 is A} \\ 0, & \text{Allele 1 is a} \end{cases}$$

$$Y = \begin{cases} 1, & \text{Allele 2 is B} \\ 0, & \text{Allele 2 is b.} \end{cases}$$

Note that $\mathrm{E}(X) = p_A$, $\mathrm{var}(X) = p_A p_a$, $\mathrm{E}(Y) = p_B$, and $\mathrm{var}(Y) = p_B p_b$. Also, following the definition of the expected value of a discrete RV and the definitions above, $\mathrm{E}(XY) = p_{AB}$:

$$\begin{aligned}
\mathrm{E}(XY) &= 1 * \mathrm{P}(X = 1 \wedge Y = 1) + 0 * \mathrm{P}(X = 0 \vee Y = 0) \\
&= \mathrm{P}(X = 1 \wedge Y = 1) \\
&= p_{AB}.
\end{aligned}$$

We now develop the value of the Pearson correlation R commonly seen in working with SNPs, $R = \frac{D}{\sqrt{p_A p_a p_B p_b}}$:

$$\begin{aligned}
R &= \frac{\mathrm{cov}(X,\ Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \\
&= \frac{\mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y)))}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \\
&= \frac{\mathrm{E}(XY) - \mathrm{E}(Y)\mathrm{E}(Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \\
&= \frac{p_{AB} - p_A p_B}{\sqrt{p_A p_a p_B p_b}} \\
&= \frac{D}{\sqrt{p_A p_a p_B p_b}}.
\end{aligned}$$