Problem Set 1 140.668 Statistical Genetics
Due 17 April 2001

1. Consider the Haldane map function, $r = (1 - e^{-2d})/2$, and the Kosambi map function, $r = \tanh(2d)/2 = (e^{4d} - 1)/[2(e^{4d} + 1)]$. (a) Find the inverse of these two functions (i.e., express d in terms of r). (b) Calculate the genetic distance (in cM) corresponding to $r = 0.01, 0.05, 0.10$ and 0.20.

2. Use Mather's formula to derive the map function for the C-L model when the distribution of the number of chiasmata on the four-strand bundle is $\boldsymbol{p} = (p_0, p_1, p_2, \dots)$. (Let $L = \sum_i ip_i/2$ denote the genetic length of the chromosome.) Use your result to show that in the case $p_n = \exp(-2L)(2L)^n/n!$, one obtains the Haldane map function.

3. Show that if the locations of chiasmata on the four-strand bundle follow a stationary Poisson process, then under no chromatid interference, the locations of crossovers on a random meiotic product also follow a stationary Poisson process.

4. Propose penetrance (emission) probabilities $e_i(v) = \Pr(\text{observe } O_i | v_i = v)$ for an intercross, to account for possible genotyping errors. Take $v \in \{\text{AA}, \text{AB}, \text{BA}, \text{BB}\}$ (phase-known genotypes) and $O \in \{\text{A}, \text{H}, \text{B}, \text{C (not AA)}, \text{D (not BB)}, \text{– (missing)}\}$.

5. Consider a nested sequence of models with $i = 0, 1, \dots, k$ markers givings $\text{RSS}_0 > \text{RSS}_1 > \dots > \text{RSS}_k$. Let $B_i(\delta) = \ln(\text{RSS}_i) + i\delta \log n/n$ (the BIC-δ criterion) and $\text{LOD}(i) = (n/2)\log_{10}(\text{RSS}_{i-1}/\text{RSS}_i)$ (a conditional LOD score). Show that picking the model for which $B_i(\delta)$ is minimized is essentially equivalent to placing a threshold on the conditional LOD score.