Problem Set 2 Special topics in genetics and genomics (140.668)
Solutions

1. The easy ones first:

$$\begin{aligned}\Pr(\text{IBS}=0|\text{IBD}=2) &= 0\\ \Pr(\text{IBS}=1|\text{IBD}=2) &= 0\\ \Pr(\text{IBS}=2|\text{IBD}=2) &= 1\end{aligned}$$

The IBD = 1 cases are pretty simple, too:

$$\begin{aligned}\Pr(\text{IBS}=0|\text{IBD}=1) &= 0\\ \Pr(\text{IBS}=1|\text{IBD}=1) &= \text{marker heterozygosity}\\ &= 1-\sum_i p_i^2\\ \Pr(\text{IBS}=2|\text{IBD}=1) &= \sum_i p_i^2\end{aligned}$$

The IBD = 0 cases are harder. We're taking two random draws from the genotypes

$$\begin{matrix}(a_1,a_1) & (a_1,a_2) & \dots & (a_1,a_k)\\ & (a_2,a_2) & \dots & (a_2,a_k)\\ & & \ddots & \vdots\\ & & & (a_k,a_k)\end{matrix}$$

with probabilities

$$\begin{matrix}p_1^2 & 2p_1p_2 & \dots & 2p_1p_k\\ & p_2^2 & \dots & 2p_2p_k\\ & & \ddots & \vdots\\ & & & p_k^2\end{matrix}$$

and we want to find the probability they share 0, 1, or 2 alleles by state.

First:

$$\begin{aligned}\Pr(\text{IBS}=2|\text{IBD}=0) &= \Pr[(11,11),(22,22),\dots,(kk,kk),(12,12),(13,13),\dots,\text{etc.})\\ &= \textstyle\sum_i p_i^4+\sum_i\sum_{j:j>i}(2p_ip_j)^2\\ \text{[now we simplify]} &= \textstyle\sum_i p_i^4+2\sum_i p_i^2\sum_{j:j\neq i}p_j^2\\ &= \textstyle\sum_i p_i^4+2\sum_i p_i^2(\sum_j p_j^2-p_i^2)\\ &= \textstyle\sum_i p_i^4+2\{(\sum_i p_i^2)^2-\sum_i p_i^4\}\\ &= \textstyle 2\{\sum_i p_i^2\}^2-\sum_i p_i^4\end{aligned}$$

The other two are similar; with (painful) simplification, we get:

$$\begin{aligned}\Pr(\text{IBS}=1|\text{IBD}=0) &= 4\{\sum p_i^2 + \sum p_i^4 - \sum p_i^3 - (\sum p_i^2)^2\} \\ \text{and } \Pr(\text{IBS}=0|\text{IBD}=0) &= 1 - 4\sum p_i^2 - 3\sum p_i^4 + 4\sum p_i^3 + 2(\sum p_i^2)^2\end{aligned}$$

Note: The biggest issues are the coefficients and the ranges of the summations (e.g., $\sum_i \sum_{j:j>i}$ or $\sum_i \sum_{j:j\neq i}$ or $\sum_i \sum_j$). *Be precise!*

Special case: $p_1 = p_2 = p_3 = p_4 = 1/4$.

	IBS		
IBD	0	1	2
0	21/64	36/64	7/64
1	0	3/4	1/4
2	0	0	1

2. Let M = parental mating type, G = kids' genotypes, and DSP = "discordant sibpair."

$$\begin{aligned}\Pr(\text{IBD}=v|\text{DSP}) &= \textstyle\sum_{M,G} \Pr(\text{IBD}=v|M,G,\text{DSP}) \Pr(G|M,\text{DSP}) \Pr(M|\text{DSP}) \\ &= \textstyle\sum_{M,G} \Pr(\text{IBD}=v|M,G) \Pr(G|M,\text{DSP}) \Pr(M|\text{DSP})\end{aligned}$$

Calculating $\Pr(M|\text{DSP})$ is similar to that for the affected sibpair, as calculated in class:

M	$\Pr(M)$	$\Pr(\text{DSP}\|M)$	$\Pr(M\|\text{DSP})$	$\Pr(M\|\text{DSP})$ when $p = 0.05$
DD × Dd	$4p^3(1-p)$	$1/2$	$\alpha \cdot 2p^3(1-p)$	~ 0.0656
Dd × Dd	$4p^2(1-p)^2$	$3/8$	$\alpha \cdot (3/2)p^2(1-p)^2$	~ 0.9344

where $\alpha = 1/[2p^3(1-p) + (3/2)p^2(1-p)^2]$

Then we calculate $\Pr(G|M,\text{DSP})$ and $\Pr(\text{IBD}=v|M,G)$ for all possible M, G:

			$\Pr(\text{IBD}=v\|M,G)$		
M	G	$\Pr(G\|M,\text{DSP})$	0	1	2
DD × Dd	DD,Dd	1	1/2	1/2	0
Dd × Dd	DD,Dd	2/3	0	1	0
Dd × Dd	DD,dd	1/3	1	0	0

We then sum up to get $\Pr(\text{IBD}=v|\text{DSP})$:

0	1	2
0.344	0.656	0

3. Let K = IBD status at disease gene and X = IBD status at marker.

$$\begin{aligned} \Pr(X = v|\text{DSP}) &= \textstyle\sum_k \Pr(X = v|\text{DSP}, K = k) \Pr(K = k|\text{DSP}) \\ &= \textstyle\sum_k \Pr(X = v|K = k) \Pr(K = k|\text{DSP}) \end{aligned}$$

For $r = 0.05$, the transition matrix $\Pr(X = x|K = k)$ is the following:

	X		
K	0	1	2
0	0.819	0.172	0.009
1	0.086	0.828	0.096
2	0.009	0.172	0.819

Note that the rows sum to 1.

We multiply each row by the result for $\Pr(K = k|\text{DSP})$ from problem 1, and obtain the following for $\Pr(X = x|\text{DSP})$.

0	1	2
0.338	0.602	0.059