Problem Set 2 solutions

140.668 Statistical Genetics

1. The inbreeding coefficient for A is the same as the kinship coefficient for her parents. Number the individuals in the pedigree from left to right, top to bottom, so that individual A's parents are 10 and 11. We'll use Wright's path method.

ancestor	path	contribution
1	$10 \cdot 7 \cdot 3 \cdot 1 \cdot 5 \cdot 8 \cdot 11$	$(1/2)^7(1+0)$
2	$10 \cdot 7 \cdot 3 \cdot 2 \cdot 5 \cdot 8 \cdot 11$	$(1/2)^7(1+0)$
4	$10 \cdot 7 \cdot 4 \cdot 8 \cdot 11$	$(1/2)^5(1+0)$
		$(1/2)^5 + 2(1/2)^7$

Thus $\Phi = 3/64 \approx 0.047$.

2. $X_i \sim \text{Binomial}(2, p_1)$, and so $\mathrm{E}X_i = 2p_i$, $\text{var}(X_i) = 2p_i(1-p_i)$.

Further,

$$\begin{aligned} \text{cov}(X_1, X_2) &= \mathrm{E}(X_1 \cdot X_2) - \mathrm{E}X_1 \cdot \mathrm{E}X_2 \\ &= \mathrm{E}(X_1 \cdot X_2) - 4p_1^2 \end{aligned}$$

We need to find the joint distribution of X_1, X_2. We condition on the IBD status of the relative pair, to find $\Pr(X_1 = i, X_2 = j | \text{IBD} = k)$, from which we obtain $\Pr(X_1 = i, X_2 = j) = \sum_{k=0}^{2} \Pr(\text{IBD} = k) \cdot \Pr(X_1 = i, X_2 = j | \text{IBD} = k)$.

IBD	X_1, X_2	$\Pr(X_1, X_2 \| \text{IBD})$
2	0,0	$(1-p_1)^2$
	1,1	$2p_1(1-p_1)$
	2,2	p_1^2
0	0,0	$(1-p_1)^4$
	0,1 (or 1,0)	$2p_1(1-p_1)^3$
	1,1	$4p_1^2(1-p_1)^2$
	0,2 (or 2,0)	$p_1^2(1-p_1)^2$
	1,2 (or 2,1)	$2p_1^3(1-p_1)$
	2,2	p_1^4
1	0,0	$(1-p_1)^3$
	0,1 (or 1,0)	$p_1(1-p_1)^2$
	1,1	$p_1(1-p_1)$
	1,2 (or 2,1)	$p_1^2(1-p_1)$
	2,2	p_1^3

From this we may obtain

$$\begin{aligned}
\mathrm{E}(X_1 \cdot X_2) &= \sum_{i=0}^{2}\sum_{j=0}^{2}\sum_{k=0}^{2} i \cdot j \cdot \Pr(\mathrm{IBD}=k) \cdot \Pr(X_1 = i, X_2 = j | \mathrm{IBD} = k) \\
&= \ldots = 4p_1^2 + p_1(1-p_1)(\pi_1 + 2\pi_2) \\
\\
\Longrightarrow \mathrm{cov}(X_1, X_2) &= p_1(1-p_1)(\pi_1 + 2\pi_2) = 4\Phi p_1(1-p_1) \\
\\
\Longrightarrow \mathrm{corr}(X_1, X_2) &= \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)\cdot\mathrm{var}(X_2)}} = 2\Phi
\end{aligned}$$

3. The easy ones first:

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

The IBD = 1 cases are pretty simple, too:

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

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

$$\begin{array}{cccc}
(a_1,a_1) & (a_1,a_2) & \ldots & (a_1,a_k) \\
 & (a_2,a_2) & \ldots & (a_2,a_k) \\
 & & \ddots & \vdots \\
 & & & (a_k,a_k)
\end{array}$$

with probabilities

$$\begin{array}{cccc}
p_1^2 & 2p_1p_2 & \ldots & 2p_1p_k \\
 & p_2^2 & \ldots & 2p_2p_k \\
 & & \ddots & \vdots \\
 & & & p_k^2
\end{array}$$

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:

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

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	0.33	0.56	0.11
1	0	0.75	0.25
2	0	0	1