## MAT 394 - PS5 Solutions

1.) Suppose that alleles $A$ and $B$ occur with frequencies 0.7 and 0.3 in a randomly-mating population at Hardy-Weinberg equilibrium.

(a) Use Bayes' theorem to find the conditional distribution of the genotypes of two individuals $M$ and $F$ given that they have a child $c1$ with genotype $AA$.

(b) Find the conditional probability that their second child, $c2$, has genotype $AB$ given that $c1$ has genotype $AA$. *Hint:* Use your results from (a) along with the law of total probability.

**Solutions:**
For (a), first observe that if we exclude mutation, then the possible genotypes for the parents are $(AA, AA)$, $(AA, AB)$, $(AB, AA)$ and $(AB, AB)$. Then, using Bayes' formula as well as the other assumptions stated in the problem, we have:

$$\begin{aligned}
\mathbb{P}(M = AA, F = AA|C_1 = AA) &= \mathbb{P}(M = AA, F = AA) \times \frac{\mathbb{P}(C_1 = AA|M = AA, F = AA)}{\mathbb{P}(C_1 = AA)} \\
&= 0.49 \times 0.49 \times \frac{1}{0.49} \\
&= 0.49 \\
\mathbb{P}(M = AA, F = AB|C_1 = AA) &= \mathbb{P}(M = AA, F = AB) \times \frac{\mathbb{P}(C_1 = AA|M = AA, F = AB)}{\mathbb{P}(C_1 = AA)} \\
&= 0.49 \times 0.42 \times \frac{0.5}{0.49} \\
&= 0.21 \\
\mathbb{P}(M = AB, F = AA|C_1 = AA) &= \mathbb{P}(M = AB, F = AA) \times \frac{\mathbb{P}(C_1 = AA|M = AB, F = AA)}{\mathbb{P}(C_1 = AA)} \\
&= 0.42 \times 0.49 \times \frac{0.5}{0.49} \\
&= 0.21 \\
\mathbb{P}(M = AB, F = AB|C_1 = AA) &= \mathbb{P}(M = AB, F = AB) \times \frac{\mathbb{P}(C_1 = AA|M = AA, F = AA)}{\mathbb{P}(C_1 = AA)} \\
&= 0.42 \times 0.42 \times \frac{1/4}{0.49} \\
&= 0.09.
\end{aligned}$$

As required, these four probabilities sum to 1. For (b), we use the law of total probability to calculate

$$\begin{aligned}
\mathbb{P}(C_2 = AB|C_1 = AA) &= \sum_{M,F} \mathbb{P}(C_2 = AB|M, F)\mathbb{P}(M, F|C_1 = AA) \\
&= \mathbb{P}(C_2 = AB|M = AA, F = AA) \times \mathbb{P}(M = AA, F = AA|C_1 = AA) + \\
&\quad \mathbb{P}(C_2 = AB|M = AA, F = AB) \times \mathbb{P}(M = AA, F = AB|C_1 = AA) + \\
&\quad \mathbb{P}(C_2 = AB|M = AB, F = AA) \times \mathbb{P}(M = AB, F = AA|C_1 = AA) + \\
&\quad \mathbb{P}(C_2 = AB|M = AB, F = AB) \times \mathbb{P}(M = AB, F = AB|C_1 = AA) \\
&= 0 \times 0.49 + 0.5 \times 0.21 + 0.5 \times 0.21 + 0.5 \times 0.09 \\
&= 0.255.
\end{aligned}$$

2.) Suppose that alleles $A$, $B$ and $C$ occur with frequencies 0.5, 0.25 and 0.25 in a randomly-mating population at Hardy-Weinberg equilibrium.

(a) Calculate the frequencies of the six possible diploid genotypes.

(b) Suppose that individuals $M$ and $F$ have genotypes $AB$ and $BC$, respectively. If they have a child $c1$, find the distribution, the mean and the variance of the number of $B$ alleles that this child inherits.

**Solutions:**

Under HWE, the genotype frequencies are $p_{AA} = 0.25$, $p_{BB} = p_{CC} = 0.0625$, $p_{AB} = p_{AC} = 0.25$, and $p_{BC} = 0.125$, which sum to 1. For (b), notice that since each parent independently transmits either 0 or 1 B alleles with equal probability, the total number transmitted to the child is binomially distributed with parameters $n = 2$ and $p = 1/2$. It follows that the expected number transmitted is $np = 1$ and the variance of the number transmitted is $np(1-p) = 1/2$.

3.) Suppose that alleles $A1$, $A2$, $A3$ and $A4$ occur with global frequencies 0.4, 0.4, 0.1 and 0.1 in a diploid population with inbreeding coefficient $\theta = 0.05$.

(a) Calculate the probability that a randomly sampled individual is a homozygote.

(b) Calculate the probability that a random sample of four chromosomes from a subpopulation will contain all four alleles.

**Solutions:**

For (a), first notice that the frequencies of the four homozygous genotypes are

$$\begin{aligned} p(A_1A_1) &= \theta 0.4 + (1-\theta)0.4^2 = 0.172 \\ p(A_2A_2) &= \theta 0.4 + (1-\theta)0.4^2 = 0.172 \\ p(A_3A_3) &= \theta 0.1 + (1-\theta)0.1^2 = 0.0145 \\ p(A_4A_4) &= \theta 0.1 + (1-\theta)0.1^2 = 0.0145. \end{aligned}$$

The homozygosity can be calculated by summing these frequencies and is equal to 0.373.

For (b), repeated use of the conditional sampling formula shows that

$$\begin{aligned} p(A_1A_2A_3A_4) &= 24 \cdot p(A_1) \cdot p_s(A_2|A_1) \cdot p_s(A_3|A_1, A_2) \cdot p_s(A_4|A_1, A_2A_3) \\ &= 24 \cdot 0.4 \cdot \frac{0 \cdot \theta + (1-\theta) \cdot 0.4}{1} \cdot \frac{0 \cdot \theta + (1-\theta) \cdot 0.1}{1+\theta} \cdot \frac{0 \cdot \theta + (1-\theta) \cdot 0.1}{1+2\theta} \\ &= 0.0285. \end{aligned}$$

Here the leading coefficient of 24 is the number of different orders that the four alleles can appear in a sample of size four.