Probability Models, solutions to Test 2

1(a) $P(X_1 = 4) = (3/4)^4 \approx 0.316$ and

$$\begin{aligned} P(X_2 = 4) &= \sum_{k=1}^{3} P(X_2 = 4 | X_1 = k) P(X_1 = k) \\ &= \sum_{k=1}^{3} \left(\frac{k}{4}\right)^4 \binom{4}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{4-k} \\ &\approx 0.147 \end{aligned}$$

Note that if $X_1 = 0$, there are no As in the first generation and that if $X_1 = 4$, A cannot *become* fixed in the second generation; it already is fixed.

(b) It is fixed in the second generation if it becomes fixed in the first or second generation. Since these are disjoint events, their probabilities can be added to give the answer $0.316 + 0.147 \approx 0.46$.

(c) $P(X_1 = 0) + P(X_1 = 4) = (1/4)^4 + (3/4)^4 \approx 0.32$

(d) With $N_1 = 1$ we get the probability $(1/4)^2 + (3/4)^2 \approx 0.62$ and with $N_1 = 4$ we get $(1/4)^8 + (3/4)^8 \approx 0.1$

2(a) If $x = y$ we get

$$M = \frac{x + x}{2} = x$$

and

$$H = \left(\frac{1}{2}\left(\frac{1}{x} + \frac{1}{x}\right)\right)^{-1} = \left(\frac{1}{x}\right)^{-1} = x$$

so that $H = M$.

(c) First note that we can write the harmonic mean as

$$H = \left(\frac{1}{2}\left(\frac{1}{x} + \frac{1}{y}\right)\right)^{-1} = \frac{2xy}{x + y}$$

where we assume that $x > 0$ and $y > 0$ so that H is well-defined. We thus need to show that

$$\frac{x+y}{2} \geq \frac{2xy}{x+y}$$

and since $x + y > 0$ this is equivalent to

$$(x+y)^2 \geq 4xy$$

Expanding the square gives

$$x^2 + y^2 \geq 2xy$$

and since

$$x^2 + y^2 - 2xy = (x-y)^2 \geq 0$$

the claim that $H \leq M$ is true. The only case of equality is when $(x-y)^2 = 0$ which is when $x = y$.

3(a) A given individual is either chosen, with probability $1/2N$, and has $2N$ offspring, or is not chosen and has 0 offspring. Hence X_k is either $2N$ with probability $1/2N$ or 0 with probability $1 - 1/2N$. We get $E[X_k] = 1$ and

$$E[X_k^2] = (2N)^2 \cdot \frac{1}{2N} + 0^2 \cdot \left(1 - \frac{1}{2N}\right) = 2N$$

which gives variance

$$\mathrm{Var}[X_k] = E[X_k^2] - (E[X_k])^2 = 2N - 1 \to \infty$$

as $N \to \infty$.

(b) In any generation, all individuals stem from the same parent and thus $T \equiv 1$.

(c) The probability a given individual is chosen is $1/2$ and hence X_k is either 2 with probability $1/2$ or 0 with probability $1/2$. We get $E[X_k] = 1$ and

$$E[X_k^2] = 2^2 \cdot \frac{1}{2} + 0^2 \cdot \frac{1}{2} = 2$$

which gives $\sigma^2 = 1$ and $E[T] \approx 4N$ generations.

(d) The probability a given individual is chosen is f and hence X_k is either $1/f$ with probability f or 0 with probability $1 - 1/f$. We then get $E[X_k] = 1$ and

$$E[X_k^2] = \left(\frac{1}{f}\right)^2 \cdot f + 0^2 \cdot \left(1 - \frac{1}{f}\right) = \frac{1}{f}$$

which gives

$$\sigma^2 = \frac{1}{f} - 1 = \frac{1-f}{f}$$

and hence

$$E[T] \approx \frac{4N}{\sigma^2} = \frac{4Nf}{1-f}$$

(e) As f increases, σ^2 decreases and we saw in Problem 3(b) on HW4 that this increases $E[T]$. Alternatively, you can argue that as f increases, the probability that a given individual is included increases. Hence, when we choose our $2N$ individuals, we expect many of them to be included and hence it becomes harder to find common ancestors.