1. Note that, as given in the notes on decoding,

$$P(\pi_3 = R \mid X) = \frac{f_R(3) \cdot g_R(3)}{P(X)}$$

To evaluate this, we can first calculate f using the forward algorithm:

$$\begin{aligned} f_V(1) &= e_V(A)(f_0(0) \cdot t_{0,V}) \\ &= 0.25(1 \cdot 0.5) = 0.125 \end{aligned}$$

$$\begin{aligned} f_R(1) &= e_R(A)(f_0(0) \cdot t_{0,R}) \\ &= 0.91(1 \cdot 0.5) = 0.455 \end{aligned}$$

$$\begin{aligned} f_V(2) &= e_V(T)(f_V(1) \cdot t_{V,V} + f_R(1) \cdot t_{R,V}) \\ &= 0.25(0.125 \cdot 0.75 + 0.455 \cdot 0.1) = 0.0348 \end{aligned}$$

$$\begin{aligned} f_R(2) &= e_R(T)(f_V(1) \cdot t_{V,R} + f_R(1) \cdot t_{R,R}) \\ &= 0.03(0.125 \cdot 0.25 + 0.455 \cdot 0.9) = 0.0132 \end{aligned}$$

$$\begin{aligned} f_V(3) &= e_V(A)(f_V(2) \cdot t_{V,V} + f_R(2) \cdot t_{R,V}) \\ &= 0.25(0.0348 \cdot 0.75 + 0.0132 \cdot 0.1) = 0.00685 \end{aligned}$$

$$\begin{aligned} f_R(3) &= e_R(A)(f_V(2) \cdot t_{V,R} + f_R(2) \cdot t_{R,R}) \\ &= 0.91(0.0348 \cdot 0.25 + 0.0132 \cdot 0.9) = 0.0185 \end{aligned}$$

$$\begin{aligned} f_V(4) &= e_V(A)(f_V(3) \cdot t_{V,V} + f_R(3) \cdot t_{R,V}) \\ &= 0.25(0.00685 \cdot 0.75 + 0.0185 \cdot 0.9) = 0.0167 \end{aligned}$$

$$\begin{aligned} f_R(4) &= e_R(A)(f_V(3) \cdot t_{V,R} + f_R(3) \cdot t_{R,R}) \\ &= 0.91(0.00685 \cdot 0.75 + 0.0185 \cdot 0.1) = 0.00175 \end{aligned}$$

$$\begin{aligned} f_V(5) &= e_V(A)(f_V(4) \cdot t_{V,V} + f_R(4) \cdot t_{R,V}) \\ &= 0.25(0.00175 \cdot 0.75 + 0.0167 \cdot 0.1) = 0.000745 \end{aligned}$$

$$\begin{aligned} f_R(5) &= e_R(A)(f_V(4) \cdot t_{V,R} + f_R(4) \cdot t_{R,R}) \\ &= 0.91(0.000438 \cdot 0.25 + 0.0167 \cdot 0.9) = 0.0141 \end{aligned}$$

And, we can calculate $g_R(3)$ using the backward algorithm:

$$g_V(5) = t_{V,E} = 1$$

$$g_R(5) = t_{R,E} = 1$$

$$\begin{aligned} g_V(4) &= t_{V,V} e_V(A) g_V(5) + t_{V,R} e_R(A) g_R(5) \\ &= (0.75 \cdot 0.25 \cdot 1) + (0.25 \cdot 0.91 \cdot 1) = 0.415 \end{aligned}$$

$$\begin{aligned} g_R(4) &= t_{R,V} e_V(A) g_V(5) + t_{R,R} e_R(A) g_R(5) \\ &= (0.1 \cdot 0.25 \cdot 1) + (0.9 \cdot 0.91 \cdot 1) = 0.844 \end{aligned}$$

$$\begin{aligned} g_V(3) &= t_{V,V} e_V(A) g_V(4) + t_{V,R} e_R(A) g_R(4) \\ &= (0.75 \cdot 0.25 \cdot 0.415) + (0.25 \cdot 0.91 \cdot 0.844) = 0.270 \end{aligned}$$

$$\begin{aligned} g_R(3) &= t_{R,V} e_V(A) g_V(4) + t_{R,R} e_R(A) g_R(4) \\ &= (0.1 \cdot 0.25 \cdot 0.415) + (0.9 \cdot 0.91 \cdot 0.844) = 0.702 \end{aligned}$$

So, we have

$$P(X) = f_V(5) + f_R(5) = 0.000745 + 0.0141 = 0.0148$$

and

$$P(\pi_3 = R \mid X) = \frac{0.0185 \cdot 0.702}{0.0148} = \mathbf{0.8775}$$

2. Given: $x_i \sim \text{Binomial}(n = 1000, p)$

   Solving for $\hat{p}$ using the **method of moments** and the fact that $\mathbb{E}_{\text{binom}}[x_i] = np$,

$$\mathbb{E}[k] = 1000p = \frac{\sum_i x_i}{10}$$

   or

$$\hat{p} = \frac{\sum_i x_i}{10000} \text{(or, 0.0385 in this particular case)}$$

   Solving for $\hat{p}$ using **least squares**:

$$\begin{aligned} \text{SSE} &= \sum_i (x_i - \mathbb{E}[x_i])^2 \\ &= \sum_i (x_i - 1000p)^2 \\ &= x_i^2 - 2000x_i p + 1000000p^2 \end{aligned}$$

   Taking the derivative with respect to p:

$$\frac{\partial}{\partial p} = -2000 \sum_i x_i + 20000000p$$

and setting this equal to 0 gives

$$\hat{p} = \frac{2000 \sum_i x_i}{20000000} = \frac{\sum_i x_i}{10000}$$

the same as the result using the method of moments.

For **maximum likelihood estimation**, note that we can write the likelihood for samples from a binomial distribution with known n as follows:

$$\begin{aligned}
L(x) &= \prod_{i=1}^{10} P_{\text{binom}}(x_i; n = 1000, p) \\
&= \prod_{i=1}^{10} \binom{1000}{x_i} p^{x_i} (1-p)^{1000-x_i} \\
&= \left[\prod_{i=1}^{10} \binom{1000}{x_i} \right] \cdot p^{\sum_i x_i} (1-p)^{10000 - \sum_i x_i}
\end{aligned}$$

Then, taking the log,

$$\ell(x) = \sum_i \log \binom{1000}{x_i} + \sum_i x_i \log p + (10000 - \sum_i x_i) \log(1-p)$$

and differentiating with respect to p:

$$\frac{\partial \ell}{\partial p} = \frac{\sum_i x_i}{p} - \frac{10000 - \sum_i x_i}{1-p}$$

Setting this to zero and multiplying both sides by $p(1-p)$ gives

$$(1-p)(\sum_i x_i) - p(10000 - \sum_i x_i) = 0$$

or,

$$\sum_i x_i - p \sum_i x_i - 10000p + p \sum_i x_i = 0$$

and canceling the $p \sum_i x_i$ and solving for p gives

$$\hat{p} = \frac{\sum_i x_i}{10000}$$

as with the previous two methods.

For **MAP estimation**, we want to estimate

$$\begin{aligned}
P(\hat{p} \mid D) &= \frac{P(D \mid \hat{p}) P(\hat{p})}{\int_{\hat{p}} P(D \mid \hat{p}) P(\hat{p})} \\
&= C \cdot \prod_{i=1}^{10} \binom{1000}{x_i} \hat{p}^{x_i} (1-\hat{p})^{1000-x_i}
\end{aligned}$$

assuming a Uniform$[0, 1]$ prior on $P(\hat{p})$, where $\frac{1}{C} = \int_{\hat{p}} P(D \mid \hat{p}) P(\hat{p})$.

This will give the same result as MLE above.

3. Note that

$$T^2 = \begin{bmatrix} 0.73 & 0.09 & 0.09 & 0.09 \\ 0.09 & 0.73 & 0.09 & 0.09 \\ 0.09 & 0.09 & 0.73 & 0.09 \\ 0.09 & 0.09 & 0.09 & 0.73 \end{bmatrix}$$

This can be used to calculate the 2^{nd} generation from the ancestor:

$$\begin{aligned} P(\text{pos } 1) = P(\text{pos } 3) = P(\text{pos } 5) \\ &= \sum_{s \in \{A,T,C,G\}} P(s \to G) \cdot P(s \to C) \cdot P(s) \\ &= 2(0.09^2 \cdot 0.25) + 2(0.73 \cdot 0.09 \cdot 0.25) \\ &= 0.0369 \end{aligned}$$

(using the symmetry of the transition matrix), and

$$\begin{aligned} P(\text{pos } 2) &= P(\text{pos } 4) \\ &= \sum_{s \in \{A,T,C,G\}} P(s \to A) \cdot P(s \to A) \cdot P(s) \\ &= (0.73^2 \cdot 0.25) + 3(0.09^2 \cdot 0.25) \\ &= 0.1392 \end{aligned}$$

so

$$\begin{aligned} P(\text{sequence } 1, \text{sequence } 2) &= (0.1392)^3 \cdot (0.0369)^2 \\ &= 3.67 \times 10^{-6} \end{aligned}$$