Question 1b :

$H_0 : \prod_{i=1}^{l}\prod_{j=1}^{9} P_{S_i N_j}$

$H_1 : \prod_{i=1}^{l} \alpha_{S_i 1} \prod_{j=1}^{8} \cdot P_{j S_{ij} S_{ij+1}}$

Using algebra, partial derivatives, and logarithms:

Exhaustive for all sites and WLOG

$L_1 = \prod_{i=1}^{l} \alpha_A^{1\{S_{i1}=A\}} \alpha_C^{1\{S_{i1}=C\}} \alpha_G^{1\{S_{i1}=G\}} \alpha_T^{1\{S_{i1}=T\}}$

$\mathrm{P}_{1AA}^{1\{S_{i2}=A \cap S_{i1}=A\}} P_{1AC}^{1\{S_{i2}=C \cap S_{i1}=A\}} P_{1AG}^{1\{S_{i2}=G \cap S_{i1}=A\}} P_{1AT}^{1\{S_{i2}=T \cap S_{i1}=A\}} ... P_{1TT}^{1\{S_{i2}=T \cap S_{i1}=T\}}$

$$l = \sum_{i=1}^{l} 1\{S_{i1} = A\} ln(\alpha_A) ... 1\{S_{i2} = A \cap S_{i1} = A\} ln(P_{1AA}) + 1\{S_{i2} = C \cap S_{i1} = A\} ln(P_{1AC})$$

$+ ... + 1\{\mathrm{S}_{i9} = T \cap S_{i8} = T\} ln(P_{8TT}) + \lambda(1 - P_{1AA} - P_{1AC} ...)$

$$\frac{\partial l_0}{\partial P_{1AA}} = \sum_{i=1}^{l} \frac{\partial}{\partial P_{1AA}} 1\{S_{i1} = A\} ln(\alpha_A) ... 1\{S_{i2} = A \cap S_{i1} = A\} ln(P_{1AA}) + 1\{S_{i2} = C \cap S_{i1} = A\} ln(P_{1AC})$$

$+ ... + 1\{\mathrm{S}_{i9} = T \cap S_{i8} = T\} ln(P_{8TT}) + \lambda(1 - P_{1AA} - P_{1AC} ...)$

$$\sum_{i=1}^{l} \frac{1\{S_{i2}=A \cap S_{i1}=A\}}{P_{1AA}} - \lambda = 0$$

$$\sum_{i=1}^{l} \frac{1\{S_{i2}=A \cap S_{i1}=A\}}{P_{1AA}} = \lambda$$

$$\frac{1}{\lambda} \sum_{i=1}^{l} 1\{S_{i2} = A \cap S_{i1} = A\} = \hat{P}_{1AA}$$

$$\frac{K_{2A|A}}{\lambda} = \hat{P}_{1AA}$$

Where $K_{2A|A}, K_{2T|A}, K_{2G|A}$ and $K_{2C|A}$ are the number of occurrences of A, T, C, and G at position 2 given the nucleotide at position 1 is an A.

Similarly taking partial derivatives for other probabilities for position 2 given position 1 is an A, and setting them equal to zero, results in $\frac{K_{2C|A}}{\lambda} = \hat{P}_{1AC}, \frac{K_{2G|A}}{\lambda} = \hat{P}_{1AG}, \frac{K_{2T|A}}{\lambda} = \hat{P}_{1AT}$

Solving for λ

$$\hat{P}_{1AA} + \hat{P}_{1AC} + \hat{P}_{1AT} + \hat{P}_{1AG} = 1$$
$$\frac{K_{2A|A}}{\lambda} + \frac{K_{2C|A}}{\lambda} + \frac{K_{2T|A}}{\lambda} + \frac{K_{2G|A}}{\lambda} = 1$$
$$K_{2A|A} + K_{2C|A} + K_{2G|A} + K_{2T|A} = \lambda$$

Therefore

$$\hat{P}_{1AA} = \frac{K_{2A|A}}{K_{2A|A} + K_{2C|A} + K_{2G|A} + K_{2T|A}}$$

Similar results can be obtained for the estimated probability at any other site $x \in \{2..9\}$ and for any combination of nucleotides $b, b'\{A, T, C, G\}$

$$\hat{P}_{xbb'} = \frac{K_{xb'|b}}{K_{xA|b} + K_{xC|b} + K_{xG|b} + K_{xT|b}}$$

$$\frac{\partial l_0}{\partial \alpha_A}\sum_{i=1}^{l} 1\{S_{i1}=A\} ln(\alpha_A)$$

$$\sum_{i=1}^{l}\frac{1\{S_{i1}=A\}}{\alpha_A}-\lambda=0$$

$$\sum_{i=1}^{l}\frac{1\{S_{i1}=A\}}{\alpha_A}=\lambda$$

$$\frac{1}{\lambda}\sum_{i=1}^{l} 1\{S_{i1}=A\}=\hat{\alpha}_A$$

$$\frac{K_{\alpha_A}}{\lambda}=\hat{\alpha}_A, \frac{K_{\alpha_C}}{\lambda}=\hat{\alpha}_C, \frac{K_{\alpha_G}}{\lambda}=\hat{\alpha}_G, \frac{K_{\alpha_T}}{\lambda}=\hat{\alpha}_T$$

$$K_{\alpha_A}+K_{\alpha_C}+K_{\alpha_G}+K_{\alpha_T}=\lambda$$

$$L_0=\prod_{i=1}^{l}\prod_{j=1}^{9} P_{iS_{ij}}^{1\{S_{ij}=A\}}...P_{iS_{ij}}^{1\{S_{ij}=G\}}$$

$$l=\sum_{i=1}^{l}\sum_{j=1}^{9} 1\{S_{ij}=A\} ln(P_{iS_{ij}})+...+1\{S_{ij}=G\} ln(P_{iS_{ij}})+\lambda(1-P_{1A}...P_{1T})$$

$$\frac{\partial l_1}{\partial p_{1A}}=\sum_{i=1}^{l}\frac{1\{S_{i1}=A\}}{P_{1A}}-\lambda=0$$

$$\sum_{i=1}^{l}\frac{1\{S_{i1}=A\}}{P_{1A}}=\lambda$$

$$\frac{1}{\lambda}\sum_{i=1}^{l} 1\{S_{i1}=A\}=\hat{P}_{1A}$$

$$\frac{K_{1A}}{\lambda}=\hat{P}_{1A}, \frac{K_{1C}}{\lambda}=\hat{P}_{1C}, \frac{K_{1G}}{\lambda}=\hat{P}_{1G}, \frac{K_{1T}}{\lambda}=\hat{P}_{1T}$$

$$\frac{K_{1A}+K_{1C}+K_{1G}+K_{1T}}{\lambda}=1$$

$$\lambda=K_{1A}+K_{1C}+K_{1G}+K_{1T}=l$$