To maximize the posterior, we need to enforce the constraint that $\sum_{j=1} \theta_{ijk} = 1$. We can do this by using a Lagrange multiplier. The constrained objective function, is given by the log likelihood plus log prior plus the constraint:

$$\ell(\boldsymbol{\theta}_{ik}, \lambda) = \sum_j N_{ijk} \log \theta_j + \lambda \left(1 - \sum_j \theta_{ijk} \right),$$

where N_{ijk} denotes $\#D \{X_i = x_{ij} \wedge Y = y_k\}$. Taking derivatives with respect to λ yields

$$\frac{\partial \ell}{\partial \lambda} = \left(1 - \sum_j \theta_{ijk} \right) = 0.$$

Taking derivatives with respect to θ_k yields

$$\begin{aligned} \frac{\partial l}{\partial \theta_{ijk}} &= \frac{N_{ijk}}{\theta_k} - \lambda = 0 \\ &\implies \theta_{ijk}\lambda = N_{ijk}. \end{aligned}$$

We can solve for λ using the sum-to-one constraint:

$$\begin{aligned} \sum_j N_{ijk} &= \lambda \sum_j \theta_{ijk} \\ N_k &= \lambda \end{aligned}.$$

where N_k denotes $\#D \{Y = y_k\}$. Thus the MAP estimate is given by

$$\hat{\theta_i}^{MAP} = \frac{N_{ijk}}{N_k} = \frac{\#D \{X_i = x_{ij} \wedge Y = y_k\}}{\#D \{Y = y_k\}},$$

which is Maximum likelihood estimates for θ_{ijk} given a set of training examples D.