MLE Derivation for a Categorical

1. $L(\theta) = \log P(C;\theta) - \lambda\left(\sum_{i=0}^{K-1} \theta_i - 1\right)$ Plug in likelihood of training corpus, apply log laws
2. $L(\theta) = \sum_i N_i \log \theta_i - \lambda(\sum_i \theta_i - 1)$
3. $\frac{\partial L}{\partial \theta_i} = \frac{N_i}{\theta_i} - \lambda = 0 \rightarrow N_i = \lambda\theta_i$ There are K of these equations.
4. $\frac{\partial L}{\partial \lambda} = -(\sum_i \theta_i - 1) = 0 \rightarrow \sum_i \theta_i = 1$
5. $\sum_i N_i = \sum_i \lambda\theta_i$ Sum up all K equations from Step 3
6. $N = \lambda$ By (4) and (5). Remember sum of all counts is N.
7. $N_i = N\theta_i$ By (3) and (6).
8. $\theta_i = \frac{N_i}{N}$