Laplacian smoothing

Idea:

- Pull some probability from frequent bigrams to infrequent ones
- Just add 1 to the counts (add-one smoothing):

$$\hat{p}(w_i|w_{i-n+1}^{i-1}) = \frac{c(w_{i-n+1}^{i}) + 1}{c(w_{i-n+1}^{i-1}) + V}$$

- Or tune a parameter (add-k smoothing):

$$\hat{p}(w_i|w_{i-n+1}^{i-1}) = \frac{c(w_{i-n+1}^{i}) + k}{c(w_{i-n+1}^{i-1}) + Vk}$$