

Improve Additive Smoothing

- Should all unseen words get equal probabilities?
- We can use a reference model to discriminate unseen words

$$p(w|d) = \begin{cases} p_{DML}(w|d) & \text{if } w \text{ is seen in } d \\ \alpha_d p(w|REF) & \text{otherwise} \end{cases}$$

Discounted ML estimate

Reference language model

$$\alpha_d = \frac{1 - \sum_{w \text{ is seen}} p_{DML}(w|d)}{\sum_{w \text{ is unseen}} p(w|REF)}$$

Normalizer
Prob. Mass for unseen words