

Smoothing Methods based on Interpolation

- Method 2: Absolute discounting (Kneser-Ney Smoothing): **Subtract a constant δ from the count of each word**

$$p(w|d) = \frac{\max(c(w,d) - \delta, 0) + \delta |d|_u p(w|REF)}{|d|}$$

unique words

- Method 3: Linear interpolation (Jelinek-Mercer smoothing): **“Shrink” uniformly toward $p(w|REF)$**

$$p(w|d) = (1 - \lambda) \frac{c(w,d)}{|d|} + \lambda p(w|REF)$$

ML estimate

parameter