

Absolute discounting

Idea:

- Let's compare the counts for bigrams in train and test sets

Experiment (Church and Gale, 1991):

- Subtract 0.75 and get a good estimate for the test count!

$$\hat{p}(w_i | w_{i-1}) = \frac{c(w_{i-1}w_i) - d}{\sum_x c(w_{i-1}x)} + \lambda(w_{i-1})p(w_i)$$