

# How to Smooth?

- All smoothing methods try to
  - discount the probability of words seen in a text data set
  - re-allocate the extra counts so that unseen words will have a non-zero count
- Method 1: Additive smoothing: **Add a constant  $\delta$  to the counts of each word, e.g., “add 1”**

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

Count of  $w$  in  $d$

“Add one”, Laplace

Vocabulary size

Length of  $d$  (total counts)