Smoothing in Naïve Bayes

- Why smoothing?
  - Address data sparseness (training data is small ➜ zero prob.)
  - Incorporate prior knowledge
  - Achieve discriminative weighting (i.e., IDF weighting)
- How?

$$p(\theta_i) = \frac{N_i + \delta}{\sum_{j=1}^{k} N_j + k\delta} \qquad \delta \geq 0$$

What if $\delta \rightarrow \infty$?

$p(w|\theta_B)$: background LM

$$p(w \mid \theta_i) = \frac{\sum_{j=1}^{N_i} c(w, d_{ij}) + \mu p(w \mid \theta_B)}{\sum_{w' \in V} \sum_{j=1}^{N_i} c(w', d_{ij}) + \mu} \qquad \mu \geq 0$$

$p(w|\theta_B)=1/|V|$?

What if $\mu \rightarrow \infty$?