

Response to Data Frequency

$d =$ text the

$$p(d|\Lambda) = [0.5 * p(\text{"text"}|\theta_d) + 0.5 * 0.1] \\ \times [0.5 * p(\text{"the"}|\theta_d) + 0.5 * 0.9]$$

$$\rightarrow p(\text{"text"}|\theta_d) = 0.9 \gg p(\text{"the"}|\theta_d) = 0.1 !$$

$d' =$ text the
the the
the ...the

$$p(d'|\Lambda) = [0.5 * p(\text{"text"}|\theta_d) + 0.5 * 0.1] \\ \times [0.5 * p(\text{"the"}|\theta_d) + 0.5 * 0.9] \\ \times [0.5 * p(\text{"the"}|\theta_d) + 0.5 * 0.9] \\ \times [0.5 * p(\text{"the"}|\theta_d) + 0.5 * 0.9]$$

...

$$\times [0.5 * p(\text{"the"}|\theta_d) + 0.5 * 0.9]$$

What if we increase $p(\theta_B)$?

What's the optimal solution now? $p(\text{"the"}|\theta_d) > 0.1$? or $p(\text{"the"}|\theta_d) < 0.1$?

Behavior 2: high frequency words get higher $p(w|\theta_d)$