Mixture of Two Unigram Language Models

- **Data**: Document d
- Mixture **Model**: **parameters** $\Lambda=(\{p(w|\theta_d)\}, \{p(w|\theta_B)\}, p(\theta_B), p(\theta_d))$
  - Two unigram LMs: θ_d **(the topic of d); θ_B (background topic)**
  - Mixing weight (topic choice): **$p(\theta_d)+p(\theta_B)=1$**
- **Likelihood** function:

$$p(d \mid \Lambda) = \prod_{i=1}^{|d|} p(x_i \mid \Lambda) = \prod_{i=1}^{|d|} [p(\theta_d)p(x_i \mid \theta_d) + p(\theta_B)p(x_i \mid \theta_B)]$$

$$= \prod_{i=1}^{M} [p(\theta_d)p(w_i \mid \theta_d) + p(\theta_B)p(w_i \mid \theta_B)]^{c(w,d)}$$

- ML **Estimate:** $\Lambda^* = \arg\max_{\Lambda} p(d \mid \Lambda)$

  Subject to $\sum_{i=1}^{M} p(w_i \mid \theta_d) = \sum_{i=1}^{M} p(w_i \mid \theta_B) = 1$ $\qquad p(\theta_d) + p(\theta_B) = 1$