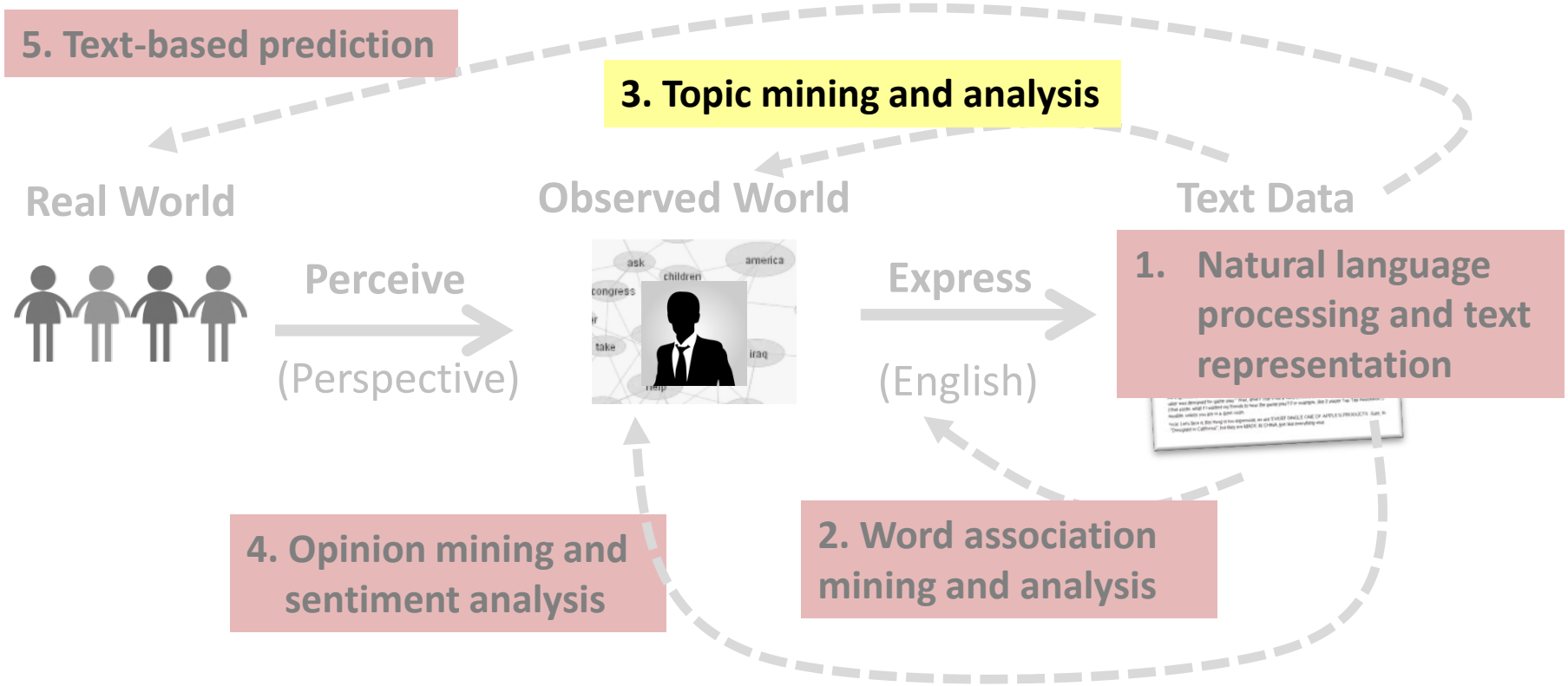
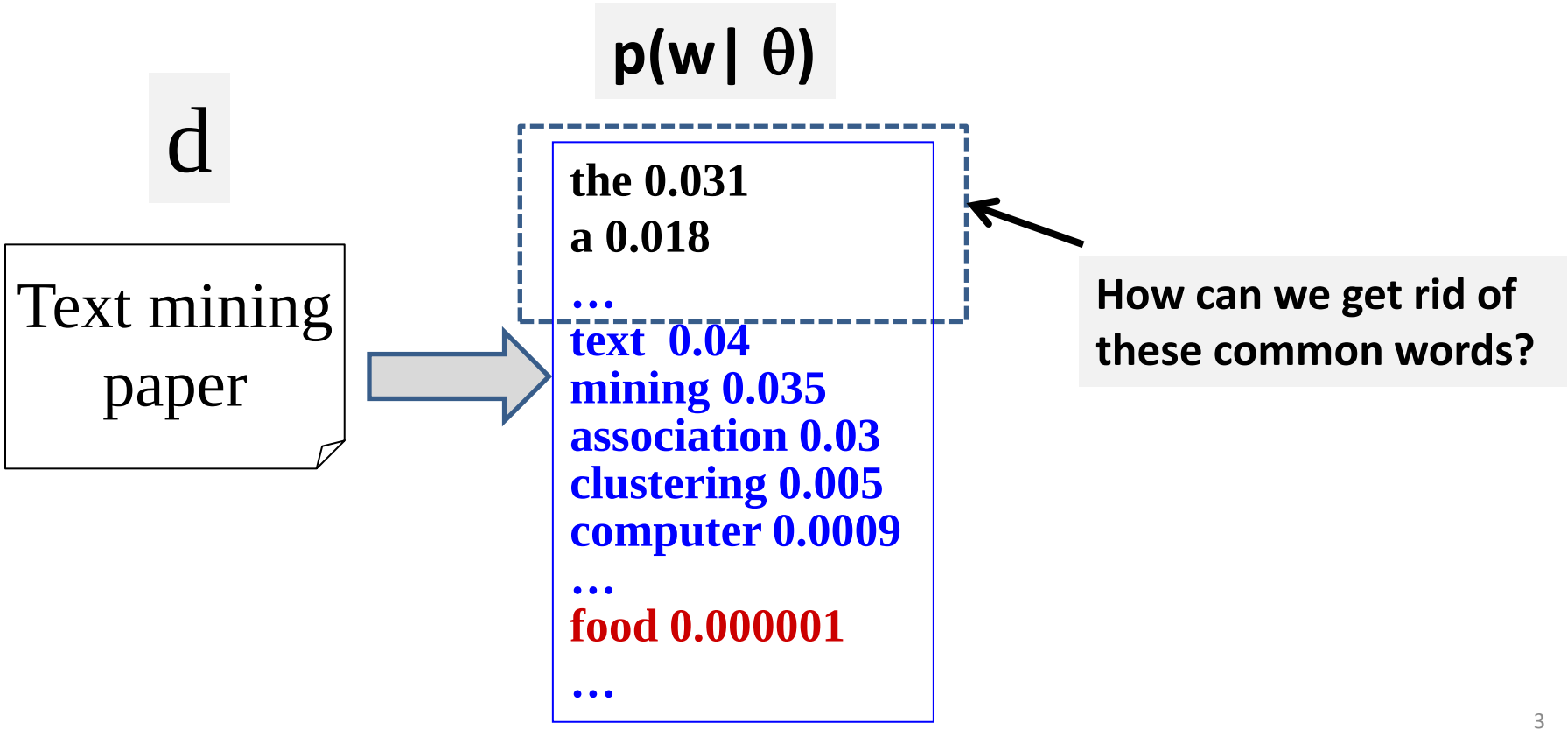
Probabilistic Topic Models: Mixture of Unigram Language Models

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

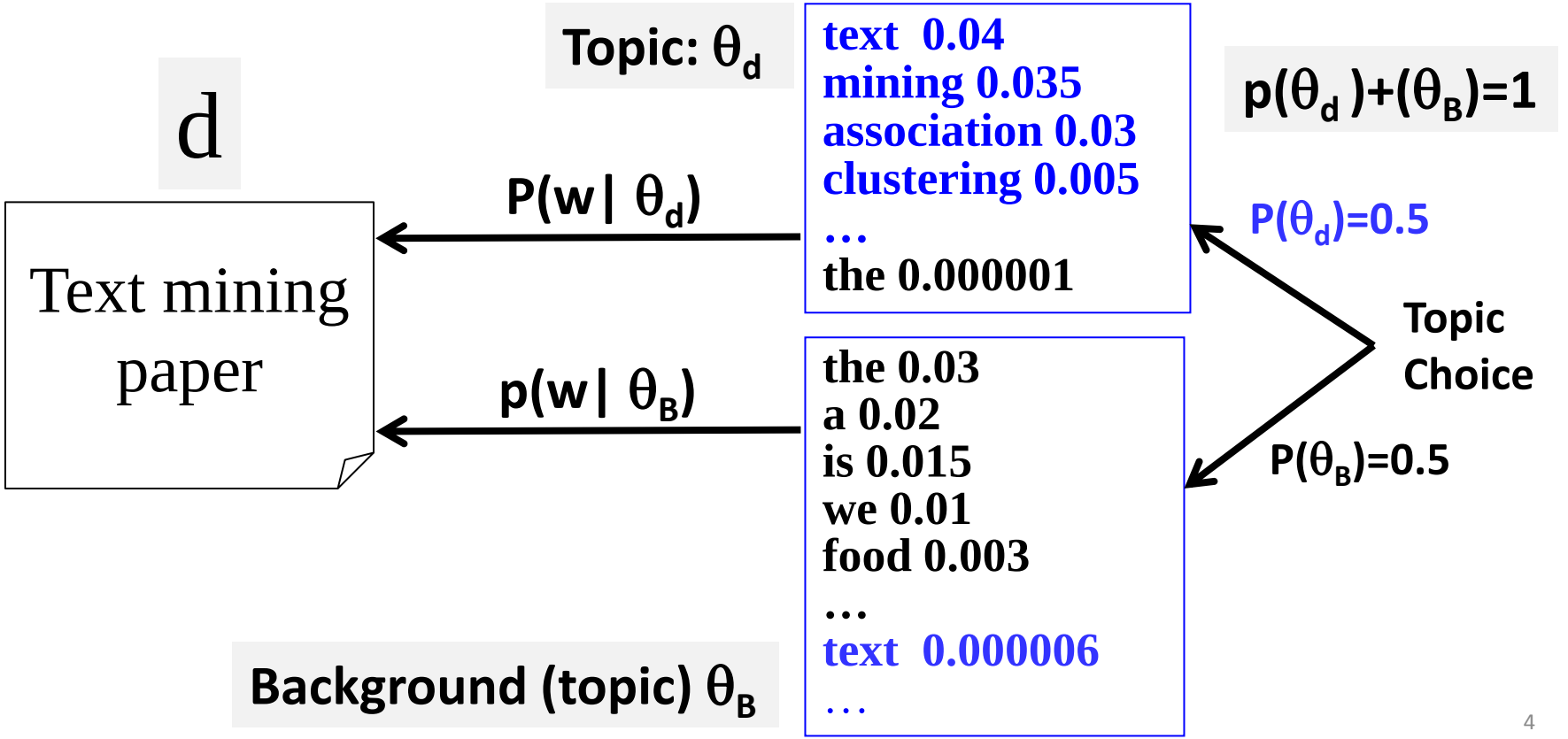
Probabilistic Topic Models: Mixture of Unigram LMs



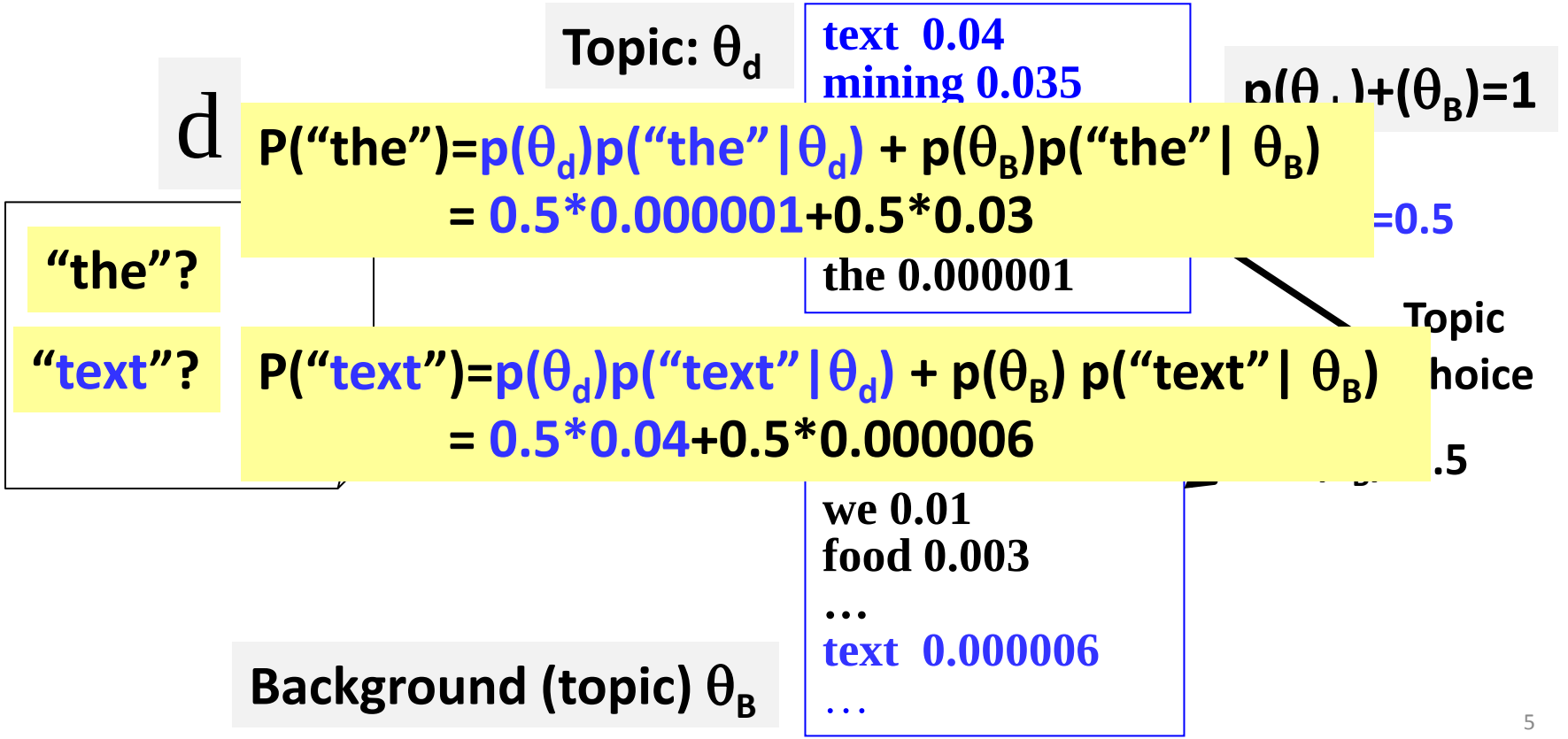
Factoring out Background Words



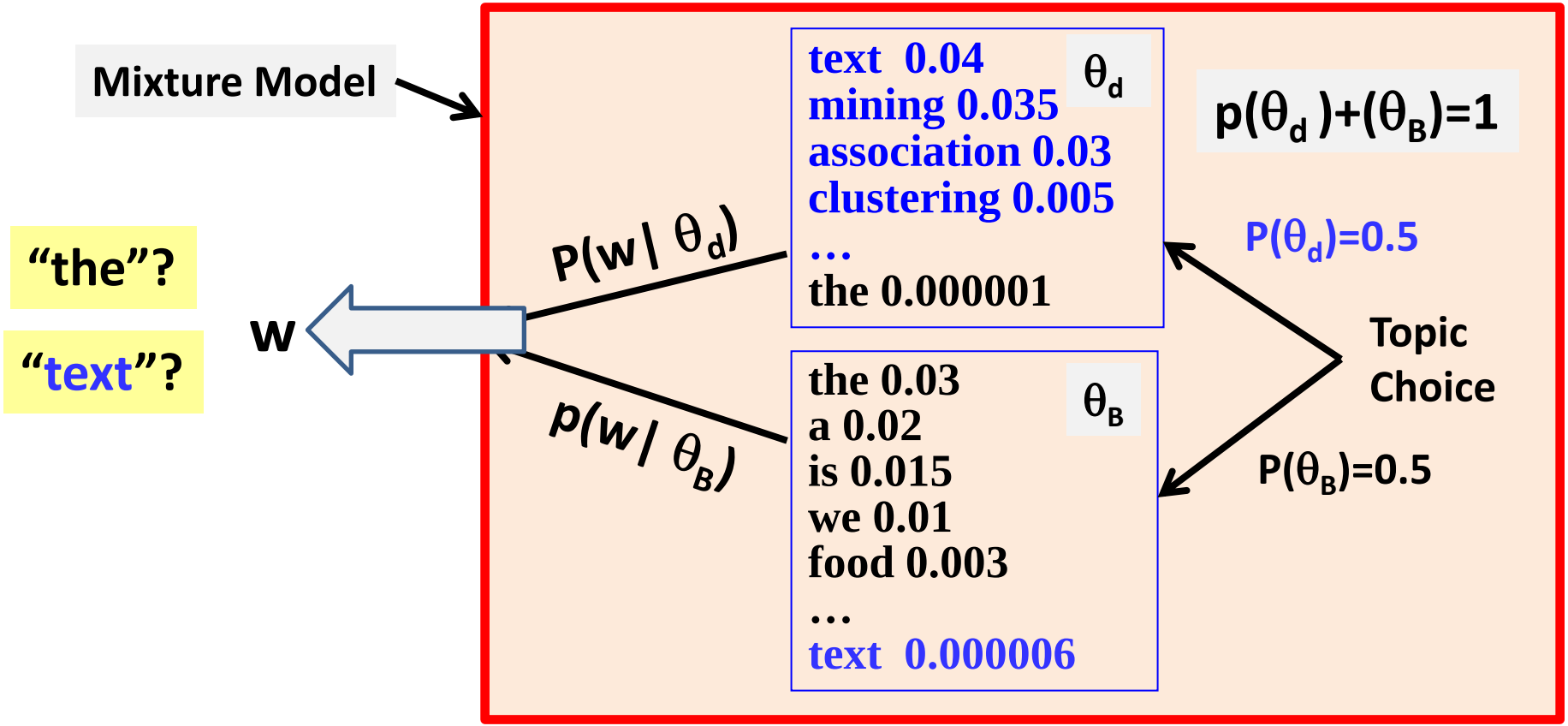
Generate d Using Two Word Distributions



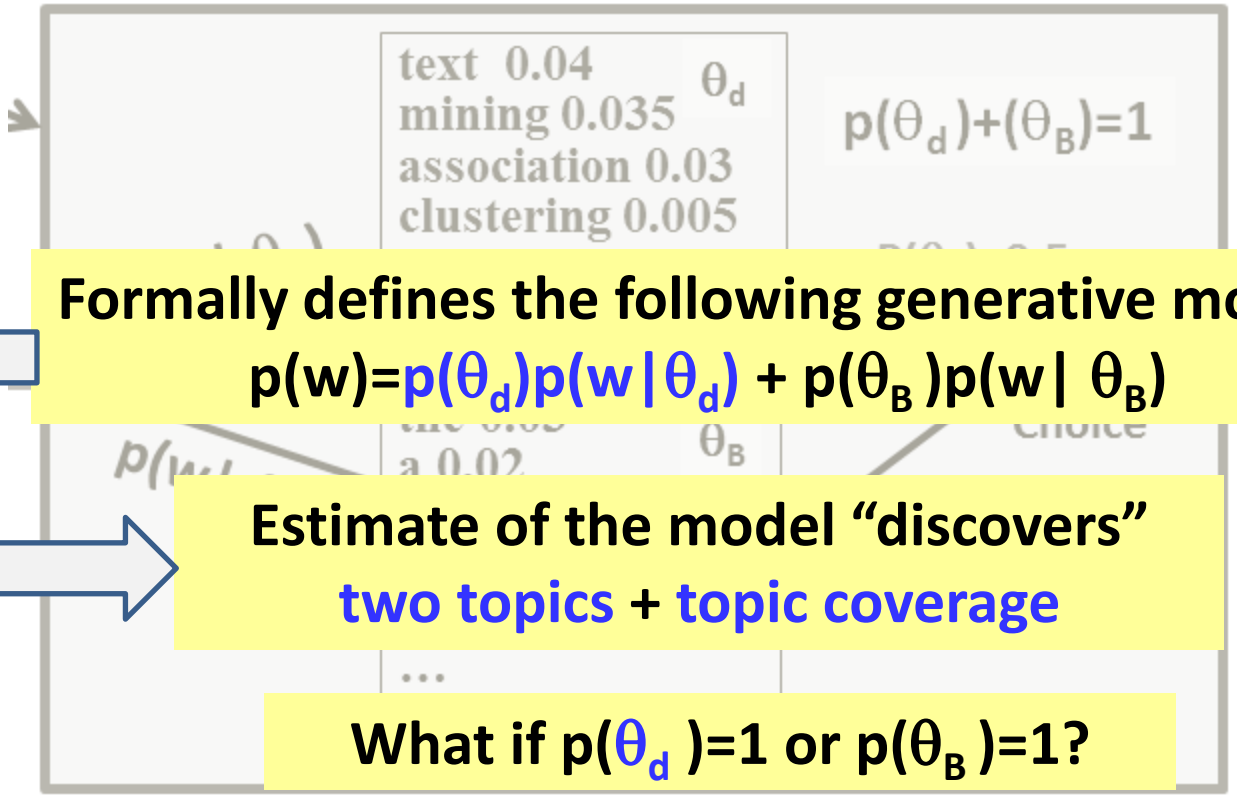
What's the probability of observing a word w?



The Idea of a Mixture Model



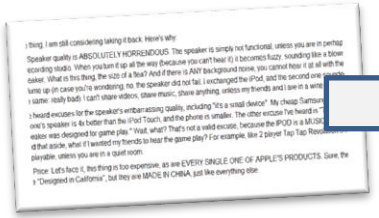
As a Generative Model…



w

Formally defines the following generative model:

$$p(w)=p(\theta_d)p(w|\theta_d) + p(\theta_B)p(w|\theta_B)$$



Estimate of the model "discovers"

two topics + topic coverage

What if $p(\theta_d)=1$ or $p(\theta_B)=1$?

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): $\mathbf{p(\theta_d)+p(\theta_B)=1}$
- **Likelihood** function:

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

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

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