

# EM Algorithm for PLSA: E-Step

**Hidden Variable (=topic indicator):  $z_{d,w} \in \{B, 1, 2, \dots, k\}$**

Probability that **w in doc d** is generated from **topic  $\theta_j$**

$$p(z_{d,w} = j) = \frac{\pi_{d,j}^{(n)} p^{(n)}(w | \theta_j)}{\sum_{j'=1}^k \pi_{d,j'}^{(n)} p^{(n)}(w | \theta_{j'})}$$

Use of Bayes Rule

$$p(z_{d,w} = B) = \frac{\lambda_B p(w | \theta_B)}{\lambda_B p(w | \theta_B) + (1 - \lambda_B) \sum_{j=1}^k \pi_{d,j}^{(n)} p^{(n)}(w | \theta_j)}$$

Probability that **w in doc d** is generated from **background  $\theta_B$**