

# EM Algorithm for PLSA: M-Step

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

Re-estimated **probability** of **doc d** covering **topic  $\theta_j$**

ML Estimate based on  
“allocated” word  
counts to topic  $\theta_j$

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w, d)(1 - p(z_{d,w} = B))p(z_{d,w} = j)}{\sum_{j'} \sum_{w \in V} c(w, d)(1 - p(z_{d,w} = B))p(z_{d,w} = j')}$$

$$p^{(n+1)}(w | \theta_j) = \frac{\sum_{d \in D} c(w, d)(1 - p(z_{d,w} = B))p(z_{d,w} = j)}{\sum_{w' \in V} \sum_{d \in D} c(w', d)(1 - p(z_{d,w'} = B))p(z_{d,w'} = j)}$$

Re-estimated **probability** of **word w** for **topic  $\theta_j$**