

Collapsed Gibbs Sampling for LDA (cont'd.)

Algorithm: for each position (j, t) , sample a new value of $z_{j,t}$ based on the current values of the rest of the z . Use the newly sampled value for computing the probability for the next sample.

Key equation:

$$P(z_{j,t} = k \mid \mathbf{Z}_{\neg j,t}, \mathbf{W}, \alpha, \beta) \propto \frac{\sigma_{j,k}^{\neg j,t} + \alpha_k}{\sum_{i=1}^K \sigma_{j,i}^{\neg j,t} + \alpha_i} \times \frac{\delta_{k,w_{j,t}}^{\neg j,t} + \beta_{w_{j,t}}}{\sum_{r=1}^V \delta_{k,r}^{\neg j,t} + \beta_r}$$

$\sigma_{j,k}$ = number of times topic k occurs in document j

$\delta_{k,v}$ = number of times word type v assigned to topic k

$\neg j, t$ = count excluding current position (j, t)