Notations

- `num_topics` $= k$ (number of topics)
- $D =$ corpus (collection of $|D|$ documents)
- `num_docs` $= |D|$ (number of documents)
- `voc_size` $= |V|$ (vocabulary size)
- `dirich_param`: an estimate of the parameter of the dirichlet distribution which generates the parameter for the (multinomial) probability distribution over topics in the document.

  size(α) = number of topics
- `word_prob_given_topic`: an array of size (nb topics, vocabulary size) which gives the (estimated) probability that a given topic will generate a certain word.

  $$\beta_{ij} = p(w^j = 1 | z^i = 1)$$

  size(β) = number of topics $\times$ vocabulary size

The following variables depend on the document d.

- `doc_size` $= |N_d|$ (number of words in the document d)
- `var_dirich`: the variational parameter for the dirichlet distribution,

  For a document d, size($\gamma^{(d)}$) = number of topics
- `var_multinom`: the variational parameter for the multinomial distribution

  For a document d, size($\phi^{(d)}$) = number of words in document d $\times$ number of topics

  $\phi^{(d)}_{ni}$ depends on the relation between the word in position n of the document and the topic i of the list of topics.
- `word_incidences`: an array containing the number of times each word in the vocabulary appears in the document.

  For a document d, size($w^{(d)}$) = number of words in document d $\times$ vocabulary size

  $w^{(d)}_{nj} = \begin{cases} 1 & \text{if the word in position } n \text{ of the document is the word in position } j \text{ of the dictionary} \\ 0 & \text{otherwise} \end{cases}$

Algorithms

Algorithm 1: E-step for a document d

Data: `word_incidences` ($w^{(d)}$), `dirich_param` (α), `word_prob_given_topic` (β)
Result: `var_dirich` ($\gamma^{(d)}$), `var_multinom` ($\phi^{(d)}$)

Algorithm 2: M-step

Data: $\{$`word_incidences` ($w^{(d)}$), `var_dirich` ($\gamma^{(d)}$), `var_multinom` ($\phi^{(d)}$), $d \in D\}$
Result: `dirich_param` (α), `word_prob_given_topic` (β)
begin

$$\beta_{ij} \propto \sum_{d \in D} \sum_{n=1}^{N_d} \phi_{ni}^{(d)} w_{nj}^{(d)}$$

Until convergence:

$$\alpha \leftarrow \alpha - H(\alpha)^{-1} g(\alpha) \text{ where } g(\alpha) = \left(\frac{\partial L}{\partial \alpha_i} \right)_i \text{ and } H(\alpha) = \left(\frac{\partial L}{\partial \alpha_i \partial \alpha_j} \right)_{ij}$$

$$\begin{cases} \dfrac{\partial L}{\partial \alpha_i} = |D| \left[\Psi\left(\displaystyle\sum_{j=1}^{k} \alpha_j \right) - \Psi(\alpha_i) \right] + \displaystyle\sum_{d \in D} \left[\Psi(\gamma_i^{(d)}) - \Psi\left(\displaystyle\sum_{j=1}^{k} \gamma_j^{(d)} \right) \right] \\ \dfrac{\partial L}{\partial \alpha_i \partial \alpha_j} = \delta_{ij} |D| \Psi'(\alpha_i) - \Psi'\left(\displaystyle\sum_{j=1}^{k} \alpha_j \right) \end{cases}$$

We note that $H(\alpha) = \operatorname{diag}(h) - 1z1^\top$ where:

$$h = (|D|\Psi'(\alpha_i))_i \quad \text{and} \quad z = \left(\Psi'\left(\sum_{p=1}^{k} \alpha_p \right) \right)_{i,j}$$

Then $(H^{-1}(\alpha)g(\alpha))_i = \dfrac{\frac{\partial L}{\partial \alpha_i} - c}{\frac{\partial^2 L}{\partial \alpha_i^2}}$ where $c = \dfrac{\sum_{j=1}^{k} \frac{\frac{\partial L}{\partial \alpha_j}}{\frac{\partial^2 L}{\partial \alpha_j^2}}}{z^{-1} + \sum_{j=1}^{k} \left(\frac{\partial^2 L}{\partial \alpha_i^2} \right)^{-1}}$.