

# Topic Model Equations

npr

March 8, 2017

The normalizing constant is the multivariate Beta function, which can be expressed in terms of the gamma function:

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}, \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K). \quad (1)$$

where  $K \geq 2$  number of categories (integer) and  $\alpha_1, \dots, \alpha_K$  concentration parameters, where  $\alpha_i > 0$ .

$$P(\mathbf{p}|\boldsymbol{\alpha}m) = \frac{\Gamma(\sum_k \alpha m_k)}{\prod_k \Gamma(\alpha m_k)} \prod_k p_k^{\alpha m_k - 1} \quad (2)$$

e.g. Slide at 13:02 in the **Topic Models** Youtube video:

$$p(\phi|\boldsymbol{\alpha}, \mathbf{w}) \propto p(\mathbf{w}|\phi) p(\phi|\boldsymbol{\alpha}) \quad (3)$$

$$\propto \prod_k \phi^{n_k} \prod_k \phi^{\alpha_k - 1} \quad (4)$$

$$\propto \prod_k \phi^{n_k + \alpha_k - 1} \quad (5)$$

$K$  topics

## 0.1 LDA Plate Notation

With plate notation, the dependencies among the many variables can be captured concisely. The boxes are plates representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.  $M$  denotes the number of documents,  $N$  the number of words in a document. Thus:

- $\boldsymbol{\alpha}$  is the parameter of the Dirichlet prior on the per-document topic distributions;

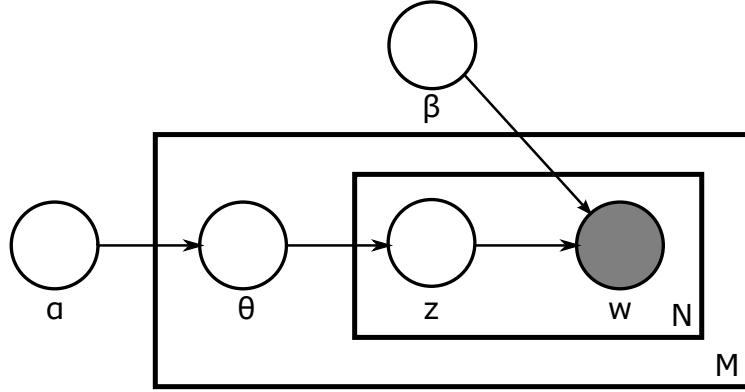


Figure 1

- $\beta$  is the parameter of the Dirichlet prior on the per-topic word distribution;
- $\theta_m$  is the topic distribution for document  $m$ ;
- $\varphi_k$  is the word distribution for topic  $k$ ;
- $z_{mn}$  is the topic for the  $n$ th word in document  $m$ , and;
- $w_{mn}$  is the specific word.

Plate notation for smoothed LDA::

The  $w_{ij}$  are the only observable variables, and the other variables are latent variables. Mostly, the basic LDA model will be extended to a smoothed version to gain better results (cite??). The plate notation is shown in Figure 1, where  $K$  denotes the number of topics considered in the model and:

$\varphi$  is a  $\mathbf{K} \times \mathbf{V}$  ( $\mathbf{V}$  is the dimension of the vocabulary) Markov matrix (transition matrix), and each row of which denotes the word distribution of a topic.

### 0.1.1 Generative Process

The generative process is as follows. Documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA assumes the following generative process for a corpus  $D$  consisting of  $M$  documents each of length  $N_i$ :

1. Choose  $\theta_i \sim \text{Dir}(\alpha)$  where  $i \in \{1, \dots, M\}$  and  $\text{Dir}(\alpha)$  is the Dirichlet distribution for parameter  $\alpha$ .
2. Choose  $\varphi_k \sim \text{Dir}(\beta)$ , where  $k \in \{1, \dots, K\}$ .

- 3. For each of the word positions  $i, j$  where  $j \in \{1, \dots, N_i\}$   $j \in \{1, \dots, N_i\}$ , and  $i \in \{1, \dots, M\}$ :
  - (a) Choose a topic  $z_{i,j} \sim \text{Multinomial}(\theta_i)$ .
  - (b) Choose a word  $w_{i,j} \sim \text{Multinomial}(\varphi_{z_{i,j}})$ . (Note that the Multinomial distribution here refers to the Multinomial with only one trial. It is formally equivalent to the categorical distribution.)

The lengths  $N_i$  are treated as independent of all the other data generating variables ( $w$  and  $z$ ). The subscript is often dropped, as in the plate diagrams shown here.