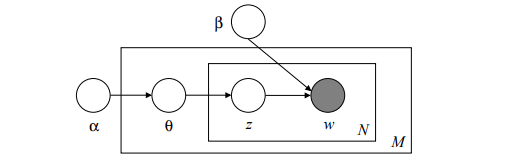
# Latent Dirichlet Allocation

## LDA过程：

1. 每篇文档的主题分布，和每个主题的词分布服从狄利克雷分布



1. 对每篇文档的N词，
2. 对每个词选取主题
3. 在主题的条件下，做



## Inference：

文档和隐藏主题标注，在先验条件和下，在已知文档主题分布和主题词分布的概率：



和的不是需要观测的变量，所以削减：



对前半部分的公式化简：



其中表示在文档m中出现主题k的次数。狄利克雷分布的归一化可表示为：



最终得到的是狄利克雷和多项分布共轨。



其中表示当前词的主题为k，则为1，否则为0。得到主题词分布的狄利克雷和多项分布共轨。

最总概率为：



## EM 方法：







estimation为假设隐藏变量的分布和在观测条件下隐藏变量分布的KL divergence。

ELBO(Evidence lower bound) 作为目标函数以及expectation。

## **Variational inference:**

